

(19) 世界知的所有権機関  
国際事務局(43) 国際公開日  
2003 年 12 月 24 日 (24.12.2003)

PCT

(10) 国際公開番号  
WO 03/107218 A1

(51) 国際特許分類: G06F 17/30, C07K 1/00

(21) 国際出願番号: PCT/JP03/06952

(22) 国際出願日: 2003 年 6 月 2 日 (02.06.2003)

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(30) 優先権データ:

特願2002-160781 2002 年 5 月 31 日 (31.05.2002) JP

特願2002-160782 2002 年 5 月 31 日 (31.05.2002) JP

特願2002-275300 2002 年 9 月 20 日 (20.09.2002) JP

特願 2002-371038

2002 年 12 月 20 日 (20.12.2002) JP

(71) 出願人 (米国を除く全ての指定国について): セレスター・レキシコ・サイエンス株式会社 (CELESTAR

LEXICO-SCIENCES, INC.) [JP/JP]; 〒261-8501 千葉県 千葉市美浜区 中瀬 1 丁目 3 番地 幕張テクノガーデン D 1 7 Chiba (JP).

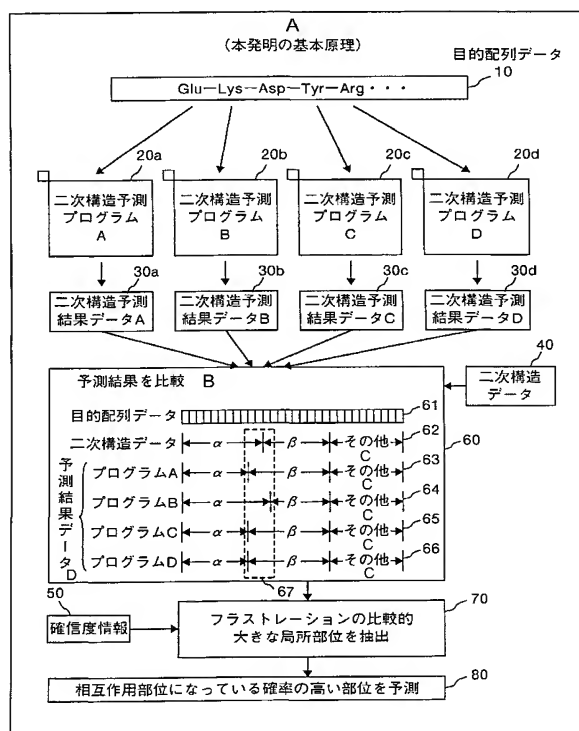
(72) 発明者; および

(75) 発明者/出願人 (米国についてのみ): 齋藤 静司 (SAITO, Seiji) [JP/JP]; 〒261-8501 千葉県 千葉市美浜区 中瀬 1 丁目 3 番地 幕張テクノガーデン D 1 7 セレスター・レキシコ・サイエンス株式会社内 Chiba (JP). 大野 一樹 (ONO, Kazuki) [JP/JP]; 〒261-8501 千葉県 千葉市美浜区 中瀬 1 丁目 3 番地 幕張テクノガーデン D 1 7 セレスター・レキシコ・サイエンス株式会社内 Chiba (JP). 和田 光人 (WADA, Mitsuhiro) [JP/JP]; 〒261-8501 千葉県 千葉市美浜区 中瀬 1 丁目 3 番地 幕張テクノガーデン D 1 7 セレスター・レキシコ・サイエンス株式会社内 Chiba (JP). 今井 建策 (IMAI, Kensaku) [JP/JP]; 〒261-8501 千葉県 千葉市美浜区 中瀬 1 丁目 3 番

[続葉有]

(54) Title: INTERACTION PREDICTING DEVICE

(54) 発明の名称: 相互作用予測装置



A... (BASIC PRINCIPLE OF THE INVENTION)

10... OBJECTIVE SEQUENCE DATA

20a... SECONDARY STRUCTURE PREDICTION PROGRAM A

20b... SECONDARY STRUCTURE PREDICTION PROGRAM B

20c... SECONDARY STRUCTURE PREDICTION PROGRAM C

20d... SECONDARY STRUCTURE PREDICTION PROGRAM D

30a... SECONDARY STRUCTURE PREDICTION RESULT DATA A

30b... SECONDARY STRUCTURE PREDICTION RESULT DATA B

30c... SECONDARY STRUCTURE PREDICTION RESULT DATA C

30d... SECONDARY STRUCTURE PREDICTION RESULT DATA D

B... COMPARE PREDICTION RESULTS

40... SECONDARY STRUCTURE DATA

61... OBJECTIVE SEQUENCE DATA

62... SECONDARY STRUCTURE DATA

C... OTHERS

D... PREDICTION RESULT DATA

63... PROGRAM A

64... PROGRAM B

65... PROGRAM C

66... PROGRAM D

50... CERTAINTY INFORMATION

70... EXTRACT LOCAL PORTION WHOSE OF FRUSTRATION IS RELATIVELY LARGE

80... PREDICT PORTION HAVING HIGH PROBABILITY THAT THE PORTION IS INTERACTION PORTION

(57) Abstract: Objective sequence data (10) which is primary sequence information on an objective protein is entered in an interaction portion predicting device by the user. A secondary structure prediction simulation of the objective sequence data (10) entered through secondary structure

[続葉有]



地 幕張テクノガーデンD17 セレスター・レキシコ・サイエンシズ株式会社内 Chiba (JP). 細木 信也 (HOSOGI,Shinya) [JP/JP]; 〒261-8501 千葉県 千葉市美浜区 中瀬 1 丁目 3 番地 幕張テクノガーデンD17 セレスター・レキシコ・サイエンシズ株式会社内 Chiba (JP). 島田 尚 (SHIMADA,Takashi) [JP/JP]; 〒261-8501 千葉県 千葉市美浜区 中瀬 1 丁目 3 番地 幕張テクノガーデンD17 セレスター・レキシコ・サイエンシズ株式会社内 Chiba (JP).

(74) 代理人: 酒井 宏明, 外(SAKAI,Hiroaki et al.); 〒100-0013 東京都千代田区霞が関三丁目2番6号 東京倶楽部ビルディング Tokyo (JP).

(81) 指定国 (国内): US.

(84) 指定国 (広域): ヨーロッパ特許 (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR).

添付公開書類:

- 国際調査報告書
- 請求の範囲の補正の期限前の公開であり、補正書受領の際には再公開される。

2文字コード及び他の略語については、定期発行される各PCTガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

prediction programs (20a to 20d) for predicting the secondary structure of the protein from the primary sequence information on the protein is executed. The results (30a to 30d) of the secondary structure predictions through the secondary structure prediction programs (20a to 20d) are compared with one another (60). On the basis of the comparison results, the frustration of a local portion of the primary sequence information of the objective protein is calculated (70). From the calculated frustration of the local portion, the interaction portion of the objective protein is predicted (80).

(57) 要約:

利用者は、目的のタンパク質の一次配列情報である目的配列データ（10）を本発明の相互作用部位予測装置に入力する。タンパク質の一次配列情報から当該タンパク質の二次構造を予測する二次構造予測プログラム（20a～20d）に対して入力された目的配列データ（10）の二次構造予測シミュレーションを実行する。各二次構造予測プログラム（20a～20d）の二次構造予測結果（30a～30d）を比較する（60）。この比較結果に基づいて、目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算する（70）。計算された局所部分のフラストレーションにより目的のタンパク質の相互作用部位を予測する（80）。

## 明 細 書

## 相互作用 予測装置

5

## 技術分野

本発明は、相互作用部位予測装置、相互作用部位予測方法、プログラム、および、記録媒体に関し、特に、局所部位のフラストレーションに基づいて相互作用部位を予測する相互作用部位予測装置、相互作用部位予測方法、プログラム、および、記録媒体に関する。

また、本発明は、活性部位予測装置、活性部位予測方法、プログラム、および、記録媒体に関し、特に、生理活性ポリペプチドもしくは蛋白質の活性部位を高精度で推定する活性部位予測装置、活性部位予測方法、プログラム、および、記録媒体に関する。

また、本発明は、タンパク質相互作用情報処理装置、タンパク質相互作用情報処理方法、プログラム、および、記録媒体に関し、特に、タンパク質の構造データから求めた疎水相互作用および静電相互作用に基づいてタンパク質単体の不安定性の高い部位を特定し相互作用部位を特定すること等のできるタンパク質相互作用情報処理装置、タンパク質相互作用情報処理方法、プログラム、および、記録媒体に関する。

また、本発明は、結合部位予測装置、結合部位予測方法、プログラム、および、記録媒体に関し、特に、アミノ酸配列データから予測された、または、実験的に求められた立体構造情報（アミノ酸残基間の空間上の距離情報）と、電荷の情報とを用いて静電的に不安定な部分を予測することにより、蛋白質または生理活性ポリペプチドの結合部位や結合相手を効率的に予測すること等がで

きる結合部位予測装置、結合部位予測方法、プログラム、および、記録媒体に関する。

さらに、本発明は、タンパク質構造最適化装置、タンパク質構造最適化方法、プログラム、および、記録媒体に関し、特に、蛋白質の構造を分割しながら所望の原子座標の最適化を行うことのできるタンパク質構造最適化装置、タンパク質構造最適化方法、プログラム、および、記録媒体に関する。

### 背景技術

(I) タンパク質が働くため、すなわちタンパク質が一定の機能を表すためには他のタンパク質や基質等となんらかの相互作用をすることが必要である。したがって、タンパク質の相互作用部位を決定することは、創薬等の分野においては極めて重要な研究テーマであり、従来より、バイオインフォマティクス分野などにおいて、タンパク質の一次配列情報（アミノ酸配列情報）に対してモチーフ検索を実行するなどの手法によりタンパク質の相互作用部位を解析する技術が開発されている。すなわち、既知の相互作用部位に特異的に存在するアミノ酸配列を探索することにより、タンパク質の相互作用部位を予測している。

しかしながら、従来のモチーフ検索などによる相互作用部位の解析手法は、既知の相互作用部位については解析することができるが、未知の相互作用部位については解析することができないというシステム構造上の基本的問題点を有していた。以下、この問題点の内容について、一層具体的に説明する。

すなわち、従来の相互作用部位の解析手法は、予め判明している相互作用部位に特異的な一次配列をモチーフデータベースなどに登録し、その情報を用いて相互作用部位の予測を行うものであるため、今までに発見されていない相互作用部位については、その解析をすることができなかった。従って、今まで発見されていない未知の相互作用部位をバイオインフォマティクス技術を用いてコンピュータ上で予測する際には、まったく別の手法を用いる必要があるが、有効な手法は未だ確立していない。



ところで、タンパク質のネイティブの立体構造は、できるだけアミノ酸間の相互作用にフラストレーションが無いような構造に形作られている。すなわち、タンパク質のエネルギー曲面はフラストレーションが無いような全体構造（ネイティブ構造）へ向かって漏斗状に設計されていると言われている（*folding funnel*）。ここで、ネイティブ構造はフラストレーションが少  
5 ない構造ではあるが、要素間に働く相互作用の複雑さ、自由度の大きさ、進化的な経緯などの点から、フラストレーションが完全に無くなっているわけではない。

最近の計算機実験では、進化の産物であるタンパク質の漏斗状のエネルギー  
10 面は本質的に等方的ではなく、フラストレーションの大きい方向と小さな方向が存在している（異方性を持つ）ことがわかってきた（*anisotropic funnel*）。これは構造的に言うならば、局所的な構造にはフラストレーションの大きい構造と小さい構造があることを示している。そして、フラストレーションの大きい局所構造部位は、全体構造の安定化のために犠牲にな  
15 っている構造部位であり、この部分は、全体構造の安定化のために歪んだ構造を取らざるを得ない状況下に置かれており、いわば全体構造の中における不安定部位であることを示している。

一方、タンパク質の相互作用は二つの安定な全体構造をもつタンパク質同士が作用することにより、さらに安定化する過程であるといえる。タンパク質が  
20 相互作用する際の構造の変化について更に説明すると、タンパク質AとBとが相互作用する際には、タンパク質Aの一部分の構造とタンパク質Bの一部分の構造に変化が起こり結合する。

変化が起こる一部分の構造として考えられる部位について考察すると、まず、局所的に見ても、かつ全体的に見ても安定な局所構造は、それ以上安定化する  
25 必要性がない。一方、全体的に見れば安定であるが局所的に見ると不安定である部分は、他のタンパク質等と結合することによりその部分が安定化し、さらに結合することにより全体もさらに安定化するという状況が考えられる。すな

わち、局所的に不安定な構造領域は、タンパク質相互作用部位である可能性が比較的高いと考えることができる。このように局所的に不安定な部位を一次配列から予測することにより、相互作用部位の候補を挙げることができるようになる可能性がある。

5       また、タンパク質の二次構造予測は、一次配列から局所的に安定な構造のパターンを予測するものである。この予測方法には、様々なものが考えられている。古くはアミノ酸の二次構造帰属情報に基づくChou-Fasmanの方法から、現在では3rd generationと呼ばれる進化的に関係のある配列をも考慮して構造予測を行う方法として、(1) neural net  
10   workを使う手法や、(2) linear statisticsを用いる手法や、(3) nearest neighbor methodを用いる手法などが考案されており、種々の異なる手法を用いて二次構造予測を行うことができる。

      これらの二次構造予測手法は、基本的には一次配列情報のうち一部の局所配列を見て予測を行うが、二次構造はタンパク質の全体の構造との関係で最終的に決定されるので、全体と局所部分との間に整合性が取れない部位、すなわち  
15   フラストレーションが大きい局所的部位では二次構造予測結果は外れる場合が多い（二次構造予測の限界）。

      ここで、このようなフラストレーションの大きい局所部位の二次構造予測に  
20   おいては、上述した様々な手法における処理の仕方の差が大きいといえるだろう。すなわち、種々の手法による誤差が大きくなる、すなわち精度の悪くなる部位等が、フラストレーションの大きい局所部位である可能性が高いといえる。よって、様々な手法における二次構造予測の結果比較を通して、フラストレーションの比較的大きな局所部位の予測が可能になると考えることができる。

25   また、タンパク質の立体構造データが分かっている場合、すなわち、既存のPDBなどに立体構造データが登録されているタンパク質は、その全体構造が既知であるため、様々な二次構造予測手法の予測結果とその実際の構造との差

を見ることにより、より明確にフラストレーションのある局所部位（相互作用部位になっている確率の高い部位）を見つけることが可能になると考えられる。

従って、本発明は、タンパク質の一次配列情報においてフラストレーションのある局所部位を発見することにより相互作用部位を効果的に予測することの  
5 できる、相互作用部位予測装置、相互作用部位予測方法、プログラム、および、記録媒体を提供することを目的としている。

（Ⅱ）また、生理活性ポリペプチド又は蛋白質の活性部位を推定する方法は種々提案されているが、それらの方法はアミノ酸配列、遺伝配列のみを用いた方法と、立体構造の情報を用いた方法の2つに大別される。

10 しかしながら、従来技術による活性部位の予測方法は予測精度が低いという問題点を有していた。

以下、この問題点の内容について、一層具体的に説明する。

まず、前者の遺伝配列のみを用いた予測方法の代表的な技術として、例えば、特開平11-213003号公報（『蛋白質機能部位の予測方法と予測装  
15 置』）のようにオリゴペプチドの出現頻度を利用して機能部位を予測する方法が挙げられる。前者の方法は、時間、計算コストの面で優れている、立体構造の情報のない蛋白質の解析に利用できる、などの利点がある。しかしながら、立体構造の情報が有る場合と比べると、精度が低いという問題点を有する。

一方、後者の立体構造を用いる活性部位予測方法において最も一般的に利用  
20 されている方法は、蛋白質の大きな溝を探す方法である。ほとんどの活性部位は結合ポケットと呼ばれる蛋白質の溝に位置している。その溝を探すことで酵素の活性部位を予測するという技術である。しかしながら、溝が複数見つかる場合や溝の位置と活性部位が一致しない場合も多く、精度が高くないという問題点を有する。さらに、活性に必要なアミノ酸残基と、単に活性部位の側に存在しているアミノ酸残基を区別することができないという問題点を有する。  
25

そこで、単にトポロジーの情報のみに頼るのではなく、計算化学を利用して、予測精度を高めようとする試みが多くされている。例えば、Ondreche

nらは活性部位においては解離性アミノ酸残基のpH滴定曲線が異常になりやすいという事実を利用して、活性部位を予測するシステムを発表している（Proc. Natl. Acad. Sci. USA, Vol. 98, Issue 22, 12473-12478, October 23, 2001）。しかし、  
5 この方法では古典論による計算を採用しているために、そもそも計算精度が低いという問題点を有する。また、引用した論文にも発表されているデータによると、pH滴定曲線が異常な解離性アミノ酸残基でも活性部位にならない場合があるという問題点を有する。

他にも、Elockらは蛋白質を不安定化させているアミノ酸残基を古典論  
10 の計算によって求めて、結合部位または活性部位になる可能性が高いことを示している（「Journal of Molecular Biology」 Vol. 312, No. 4, 885-896, September 28, 2001）。しかしながら、前述の方法と同様に古典論を利用しているために計算精度が低く、また、蛋白質を不安定化しているアミノ酸残基が、活性部位に  
15 なるという理論的な根拠はないという問題点を有する。

以上の従来技術による予測方法の問題点をまとめると、活性部位予測方法に理論的な裏付けが乏しいこと、用いている計算の精度が不十分であること等が挙げられる。これらの問題点により従来の活性部位の予測精度に限界がある。

従って、本発明は、分子軌道計算によって得られた分子軌道のエネルギーや  
20 広がりの情報から蛋白質の活性部位を予測することのできる、活性部位予測装置、活性部位予測方法、プログラム、および、記録媒体を提供することを目的としている。

（Ⅲ）また、タンパク質が働くため、すなわちタンパク質が一定の機能を表すためには他のタンパク質や基質等となんらかの相互作用をすることが必要である。  
25 したがって、タンパク質の相互作用部位を決定することは、創薬等の分野においては極めて重要な研究テーマであり、従来より、バイオインフォマティクス分野などにおいて、タンパク質の一次配列情報（アミノ酸配列情報）に対

してモチーフ検索を実行するなどの手法によりタンパク質の相互作用部位を解析する技術が開発されている。すなわち、既知の相互作用部位に特異的に存在するアミノ酸配列を探索することにより、タンパク質の相互作用部位を予測している。

- 5       しかしながら、従来のモチーフ検索などによる相互作用部位の解析手法は、既知の相互作用部位については解析することができるが、未知の相互作用部位については解析することができないというシステム構造上の基本的問題点を有していた。

- すなわち、従来の相互作用部位の解析手法は、予め判明している相互作用部位に特異的な一次配列をモチーフデータベースなどに登録し、その情報を用いて相互作用部位の予測を行うものであるため、今までに発見されていない相互作用部位については、その解析をすることができなかった。従って、今まで発見されていない未知の相互作用部位をバイオインフォマティクス技術を用いてコンピュータ上で予測する際には、まったく別の手法を用いる必要があるが、  
10       有効な手法は未だ確立していない。

- 一方、タンパク質の相互作用は二つの安定な全体構造をもつタンパク質同士が作用することにより、さらに安定化する過程であるといえる。タンパク質が相互作用する際の構造の変化について更に説明すると、タンパク質AとBとが相互作用する際には、タンパク質Aの一部分の構造とタンパク質Bの一部分の構造に変化が起こり結合する。  
20

- 変化が起こる一部分の構造として考えられる部位について考察すると、まず、局所的に見ても、かつ全体的に見ても安定な局所構造は、それ以上安定化する必要性がない。一方、全体的に見れば安定であるが局所的に見ると不安定である部分は、他のタンパク質等と結合することによりその部分が安定化し、さらに結合することにより全体もさらに安定化するという状況が考えられる。すな  
25       わち、局所的に不安定な構造領域は、タンパク質相互作用部位である可能性が比較的高いと考えることができる。このように局所的に不安定な部位を一次配

列から予測することにより、相互作用部位の候補を挙げることができるようになる可能性がある。

従って、本発明は、タンパク質の構造データから求めた疎水相互作用および静電相互作用に基づいてタンパク質単体時の不安定性の高い部位を特定し相互作用部位を特定すること等のできる、タンパク質相互作用情報処理装置、タン

5    タンパク質相互作用情報処理方法、プログラム、および、記録媒体を提供することを目的としている。

(IV) また、蛋白質または生理活性ポリペプチドが一定の機能を示すためには、他の蛋白質等と相互作用することが重要である。特定の蛋白質の相互作用を阻

10    害したり、相互作用を強めたりするような物質は医薬品になる可能性を秘めている。したがって、蛋白質の相互作用部位、さらには、蛋白質の相互作用相手を予測することは生物学、医学、薬学上、極めて重要な問題である。このため、バイオインフォマティクスの分野においては、従来から様々な方法により蛋白質の相互作用相手を予測することが試みられている。

15    しかしながら、既知のバイオインフォマティクスによる蛋白質の相互作用の予測手法は計算負担が大きく処理時間が長時間かかるものであり、また、予測精度は低いためさらなる精度の高くかつ処理時間が短い手法の開発が望まれているという問題点を有していた。

以下、この問題点の内容について、一層具体的に説明する。

20    例えば、バイオインフォマティクス分野における相互作用部位予測に関しては、モチーフ検索などによる予測技術が開発されてきた。しかし、モチーフ検索は既知の相互作用部位については解析することができるが、未知の相互作用部位については解析できないという問題点がある。

他にも、アミノ酸頻度解析を利用して結合部位を予測する方法も開発されている。例えば、特開平 1 1 - 2 1 3 0 0 3 号公報、特開平 1 0 - 2 2 2 4 8 6

25    号公報、特開平 1 0 - 0 4 5 7 9 5 号公報などがある。しかしながら、これらの予測方法は、予測精度が低いという問題点がある。

また、この他にも、例えば、2つの蛋白質の立体構造をドッキングさせ、最も安定な複合体構造を得る方法がある。この方法の予測精度は高いが、いくつかの問題点がある。第一に、蛋白質のうちに立体構造が分かっているのはほんの一部に過ぎないので、ほとんどの蛋白質には適応できないことである。第二に、これらの手法は計算負担が大きく処理時間が長時間かかるものであるため、  
5 網羅的な計算を行うことが困難であることである。

さらに、相互作用部位予測よりも難しい相互作用相手予測については有効な手段が確立されていない。つまり、全く未知の相互作用部位、さらには、相互作用相手を高精度に予測するためには、全く新しいアプローチが必要であるが、  
10 有効な手段は未だ確立していない。

従って、本発明は、バイオインフォマティクスによる蛋白質の相互作用の予測を極めて短時間で計算可能にし、網羅的な解析を可能にする、結合部位予測装置、結合部位予測方法、プログラム、および、記録媒体を提供することを目的としている。

15 (V) また、タンパク質の立体構造を基にしたドラッグデザインを行う際には、一般的に結晶構造が出発構造となる場合が多いが（例えば、「H. -D. ヘルツェ, G. フォルカース著, 江崎俊之訳, 「分子モデリング」, 地人書館, 1998年」参照。）、これには2つの問題がある。ひとつは、X線結晶回折では一般的には水素位置が決められないことである（例えば、「平山令明著,  
20 「生命科学のための結晶解析入門」, 丸善株式会社, 1996年」参照。）。欠落している水素は、モデリングソフト（例えば、Accelrys Inc. (会社名) の「WebLab Viewer Pro 4.2」(製品名) や「Insight II」(製品名) (www.accelrys.com)、Tripos, Inc. (会社名) の「SYBYL 6.7」(製品名) (www.tripos.com)、CambridgeSoft Corporation (会社名) の「Chem3D 7.0」(製品名) (www.camsoft.com) など) を用いて機械的に付加することはできるが、エネ  
25

ルギー的に安定となる配向となるわけではない。もう一つの問題は、結晶構造中では分子がパッキングされるために、いわば「干物」のような状態となっており、その構造が必ずしも生体中で機能している構造を反映しているとは限らないことである。「生身の状態」に近づけるためには、少なくとも側鎖部分を

5 リラックスさせる必要がある。そのため、局所的な原子構造の安定化のために、構造最適化が不可欠となる（例えば、「H. -D. ヘルツェ, G. フォルカー

また、タンパク質の電子状態を計算する手法には、例えば、半経験的分子軌道計算プログラムである「MOPAC 2000 ver. 1. 0」（製品名）

10 (Fujitsu Limited (会社名) Tokyo, 2000) に実装されている「MOZYME法」（例えば、「J. J. P. Stewart, Int. J. Quant. Chem., 58, 133, 1996.」参照。）などがある。この方法を用いると、2万原子程度、1000残基のタンパク質の実用的なレベルで計算できる。ただし、これは「EF (Eigenvector

15 or Following) 法」（例えば、「J. Baker, J. Comp. Chem., 7, 385, 1986.」参照。）や「BFGS (Broyden-Fletcher-Goldfarb-Shanno) 法」（例えば、「C. G. Broyden, Computr Journal, 13, 317, 1970.」、「R. Fletcher, J. Inst. Math. Appl.

20 1., 6, 222, 1970.」、「D. Goldfarb, Mathematics of Computation, 24, 23, 1970.」、「D. F. Shanno, Mathematics of Computation, 24, 647, 1970.」参照。）などの構造最適化を行わない場合である。MOPAC 2000では通常、低分子に対しては信頼性の高いEF法が用いられ、高分子に対しては収束が速く必要メモリの少量化が図れるBFGS法が用

25 いられる。

また、生体分子の計算においては溶媒効果を考慮することが重要である（例



例えば、「H. -D. ヘルツェ, G. フォルカース著, 江崎俊之訳, 「分子モデリング」, 地人書館, 1998年」および「櫻井 実, 猪飼 篤編, 「生物工学基礎コース 計算機化学入門」, 丸善株式会社, 1999」参照。 )。

5        しかしながら、蛋白質の全原子について上述したいずれかの手法により構造最適化を行う際の実用的な最適化計算は、水素原子だけ最適化する場合には800残基程度、側鎖を最適化する場合には500残基程度が限度となるというシステム構造上の問題点を有していた。

10        ここで、本問題点は近傍に位置する原子の立体障害などが主要因であるため、何も全原子を一度に計算に考慮する必要はなく、それぞれの部位について、局所的に見て安定な構造を求めてあげれば良いはずである。つまり、全体構造を部分構造に分割し、局所的な構造最適化を繰り返すことで、この問題を実用的な計算資源を用いて解決することができる。しかしながら、従来の最適化計算において、蛋白質の構造を分割しながら高精度で最適化を行う手法は存在しなかった。

15        生体分子の計算における溶媒効果の重要性について種々の文献により指摘されているが（例えば、「H. -D. ヘルツェ, G. フォルカース著, 江崎俊之訳, 「分子モデリング」, 地人書館, 1998年」および「櫻井 実, 猪飼 篤編, 「生物工学基礎コース 計算機化学入門」, 丸善株式会社, 1999」参照。）、従来の方法では溶媒効果を考慮に入れたタンパク質の構造最適化が  
20        不可能であった。

従って、本発明は、蛋白質の構造を分割しながら所望の原子座標の最適化を行うことのできる、タンパク質構造最適化装置、タンパク質構造最適化方法、プログラム、および、記録媒体を提供することを目的としている。

## 25        発明の開示

(I) 上述した目的を達成するため、本発明にかかる相互作用部位予測装置、相互作用部位予測方法、および、プログラムは、目的のタンパク質の一次配列

情報を入力する入力手段（入力ステップ）と、タンパク質の一次配列情報から当該タンパク質の二次構造を予測する二次構造予測プログラムに対して上記入力手段（入力ステップ）により入力された上記一次配列情報の二次構造予測シミュレーションを実行させる二次構造予測プログラム実行手段（二次構造予測プログラム実行ステップ）と、上記二次構造予測プログラム実行手段（二次構造予測プログラム実行ステップ）による上記二次構造予測プログラムの二次構造予測結果を比較する予測結果比較手段（予測結果比較ステップ）と、上記予測結果比較手段（予測結果比較ステップ）による比較結果に基づいて、上記目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算するフラストレーション計算手段（フラストレーション計算ステップ）と、上記フラストレーション計算手段（フラストレーション計算ステップ）により計算された上記局所部分の上記フラストレーションにより上記目的のタンパク質の相互作用部位を予測する相互作用部位予測手段（相互作用部位予測ステップ）とを備えた（含む）ことを特徴とする。

この装置、方法、および、プログラムによれば、目的のタンパク質の一次配列情報を入力し、タンパク質の一次配列情報から当該タンパク質の二次構造を予測する二次構造予測プログラムに対して入力された一次配列情報の二次構造予測シミュレーションを実行させ、二次構造予測プログラムの二次構造予測結果を比較し、比較結果に基づいて、目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算し、計算された局所部分のフラストレーションにより目的のタンパク質の相互作用部位を予測するので、タンパク質の一次配列情報においてフラストレーションのある局所部位を発見することにより相互作用部位を効果的に予測することができる。

つぎの発明にかかる相互作用部位予測装置、相互作用部位予測方法、および、プログラムは、目的のタンパク質の一次配列情報を入力する入力手段（入力ステップ）と、上記目的のタンパク質の二次構造データを取得する二次構造データ取得手段（二次構造データ取得ステップ）と、タンパク質の一次配列情報か

- ら当該タンパク質の二次構造を予測する二次構造予測プログラムに対して上記入力手段（入力ステップ）により入力された上記一次配列情報の二次構造予測シミュレーションを実行させる二次構造予測プログラム実行手段（二次構造予測プログラム実行ステップ）と、上記二次構造予測プログラム実行手段（二次構造予測プログラム実行ステップ）による上記二次構造予測プログラムの二次構造予測結果と、上記二次構造データ取得手段（二次構造データ取得ステップ）により取得した上記二次構造データとを比較する予測結果比較手段（予測結果比較ステップ）と、上記予測結果比較手段（予測結果比較ステップ）による比較結果に基づいて、上記目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算するフラストレーション計算手段（フラストレーション計算ステップ）と、上記フラストレーション計算手段（フラストレーション計算ステップ）により計算された上記局所部分の上記フラストレーションにより上記目的のタンパク質の相互作用部位を予測する相互作用部位予測手段（相互作用部位予測ステップ）とを備えた（含む）ことを特徴とする。
- この装置、方法、および、プログラムによれば、目的のタンパク質の一次配列情報を入力し、目的のタンパク質の二次構造データを取得し、タンパク質の一次配列情報から当該タンパク質の二次構造を予測する二次構造予測プログラムに対して入力された一次配列情報の二次構造予測シミュレーションを実行させ、二次構造予測プログラムの二次構造予測結果と、取得した二次構造データとを比較し、比較結果に基づいて、目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算し、計算された局所部分のフラストレーションにより目的のタンパク質の相互作用部位を予測するので、二次構造予測プログラムの予測結果と目的のタンパク質の実際の二次構造との差を見ることにより、より明確にフラストレーションのある局所部位（相互作用部位になっている確率の高い部位）を見つけることが可能になる。

つぎの発明にかかる相互作用部位予測装置、相互作用部位予測方法、および、プログラムは、上記に記載の相互作用部位予測装置、相互作用部位予測方法、

および、プログラムにおいて、上記二次構造予測プログラムの上記二次構造予測結果に対する確信度を示す確信度情報を設定する確信度情報設定手段（確信度情報設定ステップ）をさらに備え（含み）、上記フラストレーション計算手段（フラストレーション計算ステップ）は、上記確信度情報設定手段（確信度情報設定ステップ）により設定された上記確信度情報および上記比較結果に基づいて、上記局所部分の上記フラストレーションを計算することを特徴とする。

これはフラストレーション計算の一例を一層具体的に示すものである。この装置、方法、および、プログラムによれば、二次構造予測プログラムの二次構造予測結果に対する確信度を示す確信度情報を設定し、設定された確信度情報および比較結果に基づいて、局所部分のフラストレーションを計算するので、確信度情報が高い（すなわち、シミュレーションの精度の高い）プログラムによる二次構造予測結果データに対する重みを高くすることにより、フラストレーション計算においてシミュレーション結果に対する確信度を反映させることができるようになる。

また、本発明は記録媒体に関するものであり、本発明にかかる記録媒体は、上記に記載されたプログラムを記録したことを特徴とする。

この記録媒体によれば、当該記録媒体に記録されたプログラムをコンピュータに読み取らせて実行することによって、上記に記載されたプログラムをコンピュータを利用して実現することができ、これら各方法と同様の効果を得ることができる。

（Ⅱ）また、このような状況下、本発明者は、簡便でかつ正確な蛋白質の機能部位（活性部位）の推定方法等につき鋭意検討を行った結果、1）分子軌道法によって求められる蛋白質の最高被占軌道（HOMO ; highest occupied molecular orbital）または最低空軌道（LUMO ; lowest unoccupied molecular orbital）、ならびに、その周辺軌道の位置と活性部位の位置との間に関係があること、2）蛋白質の主鎖原子に分布する分子軌道の軌道エネルギーが相対

的に高いアミノ酸残基と活性部位との間に関連があることを見出し、本発明を完成した。

本発明は、以下に後述するように、1) 精度が高いと言われている分子軌道計算を利用していること、2) 福井謙一らによって提唱され、多くの科学者によって実証されてきたフロンティア軌道の位置と反応部位との関係性を蛋白質の系に応用したものであるので、理論的な根拠があるという2点によって、高い精度の予想が期待できるという特徴を有している。

すなわち、この発明の活性部位予測装置、活性部位予測方法、プログラム、および、記録媒体は、以下の考えに立脚してなされたものである。福井謙一のフロンティア軌道理論によれば、最高被占軌道 (HOMO) は化学物質の電子供与的反応に関与し、最低空軌道 (LUMO) は化学物質の電子受容的反応に関与する。この理論は低分子化合物に関してはよく実証されている。そこで、本発明者は、蛋白質のような巨大分子でも同様の理論が成立するのではないかと考えた。その可能性については計算化学的なアプローチによって提示している (Journal of the American Chemical Society; 2001; 123 (33); 8161-8162)。そして、本発明者は計算条件の改良、フロンティア軌道とその周辺の軌道という抽象的な概念から具体的な定義に変更、計算条件を詳細に検討して、さらに実施例を増やしたことにより、電子状態から逆に活性部位を予測するという本発明を完成させた。

上述した目的を達成するため、本発明にかかる活性部位予測方法は、蛋白質または生理活性ポリペプチドの電子状態を分子軌道計算によって求めて、フロンティア軌道とその周辺軌道、および／または、主鎖の重原子に局在する軌道エネルギーを特定し、当該フロンティア軌道とその周辺軌道の位置、および／または、当該軌道エネルギーに基づいて当該蛋白質または生理活性ポリペプチドの活性部位となるアミノ酸残基を予測することを特徴とする。

この方法によれば、蛋白質または生理活性ポリペプチドの電子状態を分子軌

道計算によって求めて、フロンティア軌道とその周辺軌道、および／または、主鎖の重原子に局在する軌道エネルギーを特定し、当該フロンティア軌道とその周辺軌道の位置、および／または、当該軌道エネルギーに基づいて当該蛋白質または生理活性ポリペプチドの活性部位となるアミノ酸残基を予測するので、  
5 精度が高いと言われている分子軌道計算を利用し、フロンティア軌道の位置または軌道エネルギーの高い位置と反応部位との関係性を蛋白質または生理活性ポリペプチドの系に応用するため高い精度の活性部位予想を行うことができるようになる。

つぎの発明にかかる活性部位予測装置、活性部位予測方法、および、プログラムは、目的の蛋白質または生理活性ポリペプチドの構造データを取得する構造データ取得手段（構造データ取得ステップ）と、上記構造データ取得手段（構造データ取得ステップ）にて取得した上記構造データに基づいて上記蛋白質または生理活性ポリペプチドの電子状態を分子軌道計算によって求めてフロンティア軌道を特定するフロンティア軌道計算手段（フロンティア軌道計算ステップ）と、上記フロンティア軌道と予め定めたエネルギー差になる分子軌道を当該フロンティア軌道の周辺軌道として決定する周辺軌道決定手段（周辺軌道決定ステップ）と、上記フロンティア軌道と上記周辺軌道が分布しているアミノ酸残基を活性部位の候補のアミノ酸残基として決定する候補アミノ酸残基決定手段（候補アミノ酸残基決定ステップ）と、上記候補アミノ酸残基決定手段（候補アミノ酸残基決定ステップ）により決定された上記候補アミノ酸残基の中から活性部位を選択して活性部位を予測する活性部位予測手段（活性部位予測ステップ）とを備えた（含む）ことを特徴とする。  
10  
15  
20

この装置、方法、および、プログラムによれば、目的の蛋白質または生理活性ポリペプチドの構造データを取得し、取得した構造データに基づいて蛋白質または生理活性ポリペプチドの電子状態を分子軌道計算によって求めてフロンティア軌道を特定し、フロンティア軌道と予め定めたエネルギー差になる分子軌道を当該フロンティア軌道の周辺軌道として決定し、フロンティア軌道と周  
25

5 辺軌道が分布しているアミノ酸残基を活性部位の候補のアミノ酸残基として決定し、決定された候補アミノ酸残基の中から活性部位を選択して活性部位を予測するので、精度が高いと言われている分子軌道計算を利用し、フロンティア軌道の位置と反応部位との関係性を蛋白質または生理活性ポリペプチドの系に応用するため高い精度の活性部位予想を行うことができるようになる。

つぎの発明にかかる活性部位予測装置、活性部位予測方法、および、プログラムは、目的の蛋白質または生理活性ポリペプチドの構造データを取得する構造データ取得手段（構造データ取得ステップ）と、上記構造データ取得手段（構造データ取得ステップ）にて取得した上記構造データに基づいて上記蛋白質または生理活性ポリペプチドの電子状態を分子軌道計算によって求めて主鎖の重原子に局在する軌道エネルギーを特定する軌道エネルギー計算手段（軌道エネルギー計算ステップ）と、上記軌道エネルギー計算手段（軌道エネルギー計算ステップ）にて特定された上記軌道エネルギーのうち予め定めた値を超える軌道エネルギーとなる分子軌道および／または相対的に軌道エネルギーが高い分子軌道が分布しているアミノ酸残基を活性部位の候補のアミノ酸残基として決定する候補アミノ酸残基決定手段（候補アミノ酸残基決定ステップ）とを備えた（含む）ことを特徴とする。

この装置、方法、および、プログラムによれば、目的の蛋白質または生理活性ポリペプチドの構造データを取得し、取得した構造データに基づいて蛋白質または生理活性ポリペプチドの電子状態を分子軌道計算によって求めて主鎖の重原子に局在する軌道エネルギーを特定し、特定された軌道エネルギーのうち予め定めた値を超える軌道エネルギーとなる分子軌道および／または相対的に軌道エネルギーが高い分子軌道が分布しているアミノ酸残基を活性部位の候補のアミノ酸残基として決定するので、精度が高いと言われている分子軌道計算を利用し、軌道エネルギーの高い位置と反応部位との関係性を蛋白質または生理活性ポリペプチドの系に応用するため高い精度の活性部位予想を行うことができるようになる。

つぎの発明にかかる活性部位予測装置、活性部位予測方法、および、プログラムは、目的の蛋白質または生理活性ポリペプチドの構造データを取得する構造データ取得手段（構造データ取得ステップ）と、上記構造データ取得手段（構造データ取得ステップ）にて取得した上記構造データに基づいて上記蛋白質または生理活性ポリペプチドの電子状態を分子軌道計算によって求めてフロンティア軌道を特定するフロンティア軌道計算手段（フロンティア軌道計算ステップ）と、上記構造データ取得手段（構造データ取得ステップ）にて取得した上記構造データに基づいて上記蛋白質または生理活性ポリペプチドの電子状態を分子軌道計算によって求めて主鎖の重原子に局在する軌道エネルギーを特定する軌道エネルギー計算手段（軌道エネルギー計算ステップ）と、上記フロンティア軌道と予め定めたエネルギー差になる分子軌道を当該フロンティア軌道の周辺軌道として決定する周辺軌道決定手段（周辺軌道決定ステップ）と、上記フロンティア軌道と上記周辺軌道が分布しているアミノ酸残基、および／または、上記軌道エネルギー計算手段（軌道エネルギー計算ステップ）にて特定された上記軌道エネルギーのうち予め定めた値を超える軌道エネルギーとなる分子軌道および／または相対的に軌道エネルギーが高い分子軌道が分布しているアミノ酸残基を活性部位の候補のアミノ酸残基として決定する候補アミノ酸残基決定手段（候補アミノ酸残基決定ステップ）と、上記候補アミノ酸残基決定手段（候補アミノ酸残基決定ステップ）により決定された上記候補アミノ酸残基の中から活性部位を選択して活性部位を予測する活性部位予測手段（活性部位予測ステップ）とを備えた（含む）ことを特徴とする。

この装置、方法、および、プログラムによれば、目的の蛋白質または生理活性ポリペプチドの構造データを取得し、取得した構造データに基づいて蛋白質または生理活性ポリペプチドの電子状態を分子軌道計算によって求めてフロンティア軌道を特定し、取得した構造データに基づいて蛋白質または生理活性ポリペプチドの電子状態を分子軌道計算によって求めて主鎖の重原子に局在する軌道エネルギーを特定し、フロンティア軌道と予め定めたエネルギー差になる



分子軌道を当該フロンティア軌道の周辺軌道として決定し、フロンティア軌道と周辺軌道が分布しているアミノ酸残基、および／または、特定された軌道エネルギーのうち予め定めた値を超える軌道エネルギーとなる分子軌道および／または相対的に軌道エネルギーが高い分子軌道が分布しているアミノ酸残基を  
5 活性部位の候補のアミノ酸残基として決定し、決定された候補アミノ酸残基の中から活性部位を選択して活性部位を予測するので、精度が高いといわれている分子軌道計算を利用し、フロンティア軌道の位置または軌道エネルギーの高い位置と反応部位との関係性を蛋白質または生理活性ポリペプチドの系に応用するため高い精度の活性部位予想を行うことができるようになる。

10 つぎの発明にかかる活性部位予測装置、活性部位予測方法、および、プログラムは、上記に記載の活性部位予測装置、活性部位予測方法、および、プログラムにおいて、分子軌道計算において、以下の3つの計算条件、1) 蛋白質または生理活性ポリペプチドのまわりに水分子を発生させる、2) 蛋白質または生理活性ポリペプチドの周りに連続的な誘電体を置く、3) 蛋白質または生理  
15 活性ポリペプチド表面の解離性アミノ酸残基を無電荷の状態にし、内部に埋まっている解離性アミノ酸を電荷状態にする、のうち少なくとも一つの計算条件を設定する計算条件設定手段（計算条件設定ステップ）をさらに備えた（含む）ことを特徴とする。

これは分子軌道計算の一例を一層具体的に示すものである。この装置、方法、  
20 および、プログラムによれば、分子軌道計算において、以下の3つの計算条件、1) 蛋白質または生理活性ポリペプチドのまわりに水分子を発生させる、2) 蛋白質または生理活性ポリペプチドの周りに連続的な誘電体を置く、3) 蛋白質または生理活性ポリペプチド表面の解離性アミノ酸残基を無電荷の状態にし、内部に埋まっている解離性アミノ酸を電荷状態にする、のうち少なくとも一つ  
25 の計算条件を設定するので、3つの計算条件を適宜設定することにより、効率的に分子軌道計算を実行することができ、さらに、活性部位予測の精度を大幅に向上させることができるようになる。

また、本発明は記録媒体に関するものであり、本発明にかかる記録媒体は、上記に記載されたプログラムを記録したことを特徴とする。

この記録媒体によれば、当該記録媒体に記録されたプログラムをコンピュータに読み取らせて実行することによって、上記に記載されたプログラムをコンピュータを利用して実現することができ、これら各方法と同様の効果を得ることができる。

(Ⅲ) また、上述した目的を達成するため、本発明にかかるタンパク質相互作用情報処理装置、タンパク質相互作用情報処理方法、および、プログラムは、相互作用する複数のタンパク質の一次構造データと単体時および／または複合体形成時の立体構造データとを含む構造データを取得する構造データ取得手段（構造データ取得ステップ）と、上記構造データ取得手段（構造データ取得ステップ）にて取得した上記構造データに従って、上記一次構造データを構成するアミノ酸残基毎に疎水性相互作用エネルギーを特定する疎水面特定手段（疎水面特定ステップ）と、上記構造データ取得手段（構造データ取得ステップ）にて取得した上記構造データに従って、上記一次構造データを構成するアミノ酸残基毎に静電相互作用エネルギーを特定する静電相互作用部位特定手段（静電相互作用部位特定ステップ）と、上記疎水面特定手段（疎水面特定ステップ）にて特定された上記疎水性相互作用エネルギー、および、上記静電相互作用部位特定手段（静電相互作用部位特定ステップ）にて特定された上記静電相互作用エネルギーに基づいて不安定性の高い上記アミノ酸残基の部位を特定することにより相互作用部位を特定する相互作用部位特定手段（相互作用部位特定ステップ）とを備えた（含む）ことを特徴とする。

この装置、方法、および、プログラムによれば、相互作用する複数のタンパク質の一次構造データと単体時および／または複合体形成時の立体構造データとを含む構造データを取得し、取得した構造データに従って、一次構造データを構成するアミノ酸残基毎に疎水性相互作用エネルギーを特定し、取得した構造データに従って、一次構造データを構成するアミノ酸残基毎に静電相互作用

エネルギーを特定し、特定された疎水性相互作用エネルギー、および、静電相互作用エネルギーに基づいて不安定性の高いアミノ酸残基の部位を特定することにより相互作用部位を特定するので、構造データによりタンパク質の相互作用部位を容易に特定することができるようになる。

- 5 つぎの発明にかかるタンパク質相互作用情報処理装置、タンパク質相互作用情報処理方法、および、プログラムは、上記に記載のタンパク質相互作用情報処理装置、タンパク質相互作用情報処理方法、および、プログラムにおいて、上記構造データ取得手段（構造データ取得ステップ）にて取得した上記構造データに従って、上記一次構造データを構成するアミノ酸残基毎に溶媒接触面を
- 10 特定する溶媒接触面特定手段（溶媒接触面特定ステップ）をさらに備え（含み）、上記相互作用部位特定手段（相互作用部位特定ステップ）は、上記溶媒接触面特定手段（溶媒接触面特定ステップ）にて特定された上記溶媒接触面、上記疎水面特定手段（疎水面特定ステップ）にて特定された上記疎水性相互作用エネルギー、および、上記静電相互作用部位特定手段（静電相互作用部位特定
- 15 定ステップ）にて特定された上記静電相互作用エネルギーに基づいて不安定性の高い上記アミノ酸残基の部位を特定することにより相互作用部位を特定することを特徴とする。

- この装置、方法、および、プログラムによれば、取得した構造データに従って、一次構造データを構成するアミノ酸残基毎に溶媒接触面を特定し、特定
- 20 れた溶媒接触面、疎水性相互作用エネルギー、および、静電相互作用エネルギーに基づいて不安定性の高いアミノ酸残基の部位を特定することにより相互作用部位を特定するので、複合体時の構造データがある場合に、タンパク質の相互作用部位をさらに正確かつ容易に特定することができるようになる。

- つぎの発明にかかるタンパク質相互作用情報処理装置、タンパク質相互作用
- 25 情報処理方法、および、プログラムは、上記に記載のタンパク質相互作用情報処理装置、タンパク質相互作用情報処理方法、および、プログラムにおいて、上記相互作用部位特定手段（相互作用部位特定ステップ）にて特定された上記

相互作用部位について、相互作用する相手側の一次配列を特定し、当該一次配列を含む一次構造を持つ候補タンパク質を検索する候補タンパク質検索手段

(候補タンパク質検索ステップ) をさらに備え(含み)、上記候補タンパク質検索手段(候補タンパク質検索ステップ) にて検索された上記候補タンパク質

- 5 について、上記相手側の一次配列の部分が上記候補タンパク質の相互作用部位として特定されるか否かを確認することを特徴とする。

- この装置、方法、および、プログラムによれば、相互作用部位特定手段(相互作用部位特定ステップ) にて特定された相互作用部位について、相互作用する相手側の一次配列を特定し、当該一次配列を含む一次構造を持つ候補タンパク質を検索し、検索された候補タンパク質について、上述した構造データ取得手段(構造データ取得ステップ)、溶媒接触面特定手段(溶媒接触面特定ステップ)(複合体時の構造データがある場合)、疎水面特定手段(疎水面特定ステップ)、静電相互作用部位特定手段(静電相互作用部位特定ステップ)、および、相互作用部位特定手段(相互作用部位特定ステップ) を実行することにより、相手側の一次配列の部分が候補タンパク質の相互作用部位として特定されるか否かを確認するので、未知の相互作用を予測することが容易にできるようになる。
- 10
- 15

また、本発明は記録媒体に関するものであり、本発明にかかる記録媒体は、上記に記載されたプログラムを記録したことを特徴とする。

- 20 この記録媒体によれば、当該記録媒体に記録されたプログラムをコンピュータに読み取らせて実行することによって、上記に記載されたプログラムをコンピュータを利用して実現することができ、これら各方法と同様の効果を得ることができる。

- (IV) また、2つの蛋白質が自発的に相互作用するためには、結合することにより系全体のエネルギーが下がることが必要となる。つまり、蛋白質の中で不安定な部分は、結合によって安定化する可能性を秘めているので、結合しやすい部分と考えることができる。また、相互作用相手は他の蛋白質と比べて結合
- 25

する能力が高いと考えられる。つまり、相互作用相手を予測するためには、網羅的に相互作用を計算した上で、他よりも相互作用する能力が高いものを探索する必要がある。そのためには、1対1の相互作用だけではなく、多対多の相互作用の計算が必要なので、計算コストを大幅に改善する必要がある。

- 5      本発明の中心的概念は、蛋白質の構造上、他の部位よりも不安定な領域が結合部位の可能性が高いというものである。つまり、本発明は、局所的に不安定な領域を比較的簡易な計算によって求めることで、結合部位を予測するものである。

すなわち、本発明は、主に、基本的に蛋白質の配列情報のみから高精度に結合部位を予測できるようにすること（ただし、必要に応じて立体構造情報を追加できる。）極めて短時間で計算可能にし、網羅的な解析を可能にすること、の2点を特徴とするものである。

従って、本発明は、蛋白質のアミノ酸情報から立体構造情報（アミノ酸間の空間上の距離）を予測して、立体構造と電荷の情報から静電的に不安定な部分を予測する、および／または、2つの蛋白質が結合する時の静電的なエネルギーを計算することにより、結合部位、結合相手を予測すること等ができる、結合部位予測装置、結合部位予測方法、プログラム、および、記録媒体に関する。

上述した目的を達成するため、本発明にかかる結合部位予測方法は、蛋白質または生理活性ポリペプチドのアミノ酸配列データから当該蛋白質または生理活性ポリペプチドの立体構造における各アミノ酸残基間の空間上の距離データを求め、当該距離データと各アミノ酸の電荷に従って静電的に不安定なアミノ酸残基を特定することにより結合部位を予測することを特徴とする。

この方法によれば、蛋白質または生理活性ポリペプチドのアミノ酸配列データから当該蛋白質または生理活性ポリペプチドの立体構造における各アミノ酸残基間の空間上の距離データを求め、当該距離データと各アミノ酸の電荷に従って静電的に不安定なアミノ酸残基を特定することにより結合部位を予測するので、蛋白質または生理活性ポリペプチドのアミノ酸配列から静電的に不安定

になりそうなアミノ酸残基が結合部位になりやすいことを利用して、高速かつ高精度に結合部位を予測することができるようになる。

つぎの発明にかかる結合部位予測装置、結合部位予測方法、および、プログラムは、目的の蛋白質または生理活性ポリペプチドのアミノ酸配列データを取得するアミノ酸配列データ取得手段（アミノ酸配列データ取得ステップ）と、  
5 上記アミノ酸配列データ取得手段（アミノ酸配列データ取得ステップ）により取得されたアミノ酸配列データに含まれる各アミノ酸残基の間の空間上の距離を決定する空間距離決定手段（空間距離決定ステップ）と、上記アミノ酸配列データに含まれる各アミノ酸残基が持つ電荷を決定する電荷決定手段（電荷決定ステップ）と、上記空間距離決定手段（空間距離決定ステップ）により決定された各アミノ酸残基間の空間上の距離と、上記電荷決定手段（電荷決定ステップ）により決定された各アミノ酸残基が持つ電荷とに従って、各アミノ酸残基のエネルギーを計算するエネルギー計算手段（エネルギー計算ステップ）と、  
10 上記エネルギー計算手段（エネルギー計算ステップ）により計算された上記エネルギーに従って、結合部位となる候補アミノ酸残基を決定する候補アミノ酸残基決定手段（候補アミノ酸残基決定ステップ）とを備えた（含む）ことを特徴とする。

この装置、方法、および、プログラムによれば、目的の蛋白質または生理活性ポリペプチドのアミノ酸配列データを取得し、取得されたアミノ酸配列データに含まれる各アミノ酸残基の間の空間上の距離を決定し、アミノ酸配列データに含まれる各アミノ酸残基が持つ電荷を決定し、決定された各アミノ酸残基間の空間上の距離と、決定された各アミノ酸残基が持つ電荷とに従って、各アミノ酸残基のエネルギーを計算し、計算されたエネルギーに従って、結合部位となる候補アミノ酸残基を決定するので、蛋白質または生理活性ポリペプチド  
20 のアミノ酸配列から静電的に不安定になりそうなアミノ酸残基が結合部位になりやすいことを利用して、高速かつ高精度に結合部位を予測することができるようになる。

つぎの発明にかかる結合部位予測装置、結合部位予測方法、および、プログラムは、目的の複数の蛋白質または生理活性ポリペプチドのアミノ酸配列データを取得するアミノ酸配列データ取得手段（アミノ酸配列データ取得ステップ）と、上記目的の複数の蛋白質または生理活性ポリペプチドの結合した複合体の立体構造情報を生成する複合体構造生成手段（複合体構造生成ステップ）と、上記アミノ酸配列データ取得手段（アミノ酸配列データ取得ステップ）により取得された複数のアミノ酸配列データに含まれる各アミノ酸残基の間の空間上の距離を、上記複合体構造生成手段（複合体構造生成ステップ）により生成した上記複合体の立体構造情報に従って決定する空間距離決定手段（空間距離決定ステップ）と、上記複数のアミノ酸配列データに含まれる各アミノ酸残基が持つ電荷を決定する電荷決定手段（電荷決定ステップ）と、上記空間距離決定手段（空間距離決定ステップ）により決定された各アミノ酸残基間の空間上の距離と、上記電荷決定手段（電荷決定ステップ）により決定された各アミノ酸残基が持つ電荷とに従って、各アミノ酸残基のエネルギーを計算するエネルギー計算手段（エネルギー計算ステップ）と、上記複合体構造生成手段（複合体構造生成ステップ）により上記複合体について結合部位を変えて上記複合体の立体構造情報を生成し、上記エネルギー計算手段（エネルギー計算ステップ）により各アミノ酸残基のエネルギーを計算し、当該エネルギーの総和が最小となる結合部位を求めるエネルギー最小化手段（エネルギー最小化ステップ）と、上記エネルギー最小化手段（エネルギー最小化ステップ）によりエネルギーの総和が最小となる結合部位を、結合部位の候補アミノ酸残基として決定する候補アミノ酸残基決定手段（候補アミノ酸残基決定ステップ）とを備えた（含む）ことを特徴とする。

この装置、方法、および、プログラムによれば、目的の複数の蛋白質または生理活性ポリペプチドのアミノ酸配列データを取得し、目的の複数の蛋白質または生理活性ポリペプチドの結合した複合体の立体構造情報を生成し、取得された複数のアミノ酸配列データに含まれる各アミノ酸残基の間の空間上の距離

を、生成した複合体の立体構造情報に従って決定し、複数のアミノ酸配列データに含まれる各アミノ酸残基が持つ電荷を決定し、決定された各アミノ酸残基間の空間上の距離と、決定された各アミノ酸残基が持つ電荷とに従って、各アミノ酸残基のエネルギーを計算し、複合体について結合部位を変えて複合体の  
5 立体構造情報を生成し、各アミノ酸残基のエネルギーを計算し、当該エネルギーの総和が最小となる結合部位を求め、エネルギーの総和が最小となる結合部位を、結合部位の候補アミノ酸残基として決定するので、蛋白質または生理活性ポリペプチドのアミノ酸配列から静電的に不安定になりそうなアミノ酸残基が結合部位になりやすいことを利用して、高速かつ高精度に結合部位を予測す  
10 ることができるようになる。

つぎの発明にかかる結合部位予測装置、結合部位予測方法、および、プログラムは、目的の蛋白質または生理活性ポリペプチドのアミノ酸配列データと、結合候補となる1つまたは複数の蛋白質または生理活性ポリペプチドのアミノ酸配列データとを取得するアミノ酸配列データ取得手段（アミノ酸配列データ  
15 取得ステップ）と、上記目的の蛋白質または生理活性ポリペプチドと、上記結合候補となる蛋白質または生理活性ポリペプチドとが結合した複合体の立体構造情報を生成する複合体構造生成手段（複合体構造生成ステップ）と、上記アミノ酸配列データ取得手段（アミノ酸配列データ取得ステップ）により取得された目的のアミノ酸配列データと結合候補のアミノ酸配列データに含まれる各  
20 アミノ酸残基の間の空間上の距離を、上記複合体構造生成手段（複合体構造生成ステップ）により生成した上記複合体の立体構造情報に従って決定する空間距離決定手段（空間距離決定ステップ）と、上記目的のアミノ酸配列データと上記結合候補のアミノ酸配列データに含まれる各アミノ酸残基が持つ電荷を決定する電荷決定手段（電荷決定ステップ）と、上記空間距離決定手段（空間距  
25 離決定ステップ）により決定された各アミノ酸残基間の空間上の距離と、上記電荷決定手段（電荷決定ステップ）により決定された各アミノ酸残基が持つ電荷とに従って、各アミノ酸残基のエネルギーを計算するエネルギー計算手段



- (エネルギー計算ステップ) と、上記複合体構造生成手段 (複合体構造生成ステップ) により上記複合体について結合部位を変えて上記複合体の立体構造情報を生成し、上記エネルギー計算手段 (エネルギー計算ステップ) により各アミノ酸残基のエネルギーを計算し、当該エネルギーの総和が最小となる結合部位を求め、上記エネルギー最小化手段 (エネルギー最小化ステップ) と、全ての結合候補について上記エネルギー最小化手段 (エネルギー最小化ステップ) を実行した結果、上記エネルギーの総和が最小となる結合部位をもつ結合候補を決定する結合候補決定手段 (結合候補決定ステップ) とを備えた (含む) ことを特徴とする。
- 5
- 10 この装置、方法、および、プログラムによれば、目的の蛋白質または生理活性ポリペプチドのアミノ酸配列データと、結合候補となる 1 つまたは複数の蛋白質または生理活性ポリペプチドのアミノ酸配列データとを取得し、目的の蛋白質または生理活性ポリペプチドと、結合候補となる蛋白質または生理活性ポリペプチドとが結合した複合体の立体構造情報を生成し、取得された目的のア
- 15 ミノ酸配列データと結合候補のアミノ酸配列データに含まれる各アミノ酸残基の間の空間上の距離を、生成した複合体の立体構造情報に従って決定し、目的のアミノ酸配列データと結合候補のアミノ酸配列データに含まれる各アミノ酸残基が持つ電荷を決定し、決定された各アミノ酸残基間の空間上の距離と、決定された各アミノ酸残基が持つ電荷とに従って、各アミノ酸残基のエネルギー
- 20 を計算し、複合体について結合部位を変えて複合体の立体構造情報を生成し、各アミノ酸残基のエネルギーを計算し、当該エネルギーの総和が最小となる結合部位を求め、全ての結合候補についてエネルギー最小化処理を実行した結果、エネルギーの総和が最小となる結合部位をもつ結合候補を決定するので、蛋白質または生理活性ポリペプチドのアミノ酸配列から静電的に不安定になりそう
- 25 なアミノ酸残基が結合部位になりやすいことを利用して、高速かつ高精度に最適な結合候補の蛋白質を予測することができるようになる。

また、本発明は記録媒体に関するものであり、本発明にかかる記録媒体は、

上記に記載されたプログラムを記録したことを特徴とする。

この記録媒体によれば、当該記録媒体に記録されたプログラムをコンピュータに読み取らせて実行することによって、上記に記載されたプログラムをコンピュータを利用して実現することができ、これら各方法と同様の効果を得ることができる。

- (V) また、上述した目的を達成するため、本発明にかかるタンパク質構造最適化装置、タンパク質構造最適化方法、および、プログラムは、タンパク質の座標データを取得する座標データ取得手段（座標データ取得ステップ）と、上記タンパク質の座標データについて、特定のアミノ酸残基から所定の距離内に含まれる近傍アミノ酸残基群の座標を抽出する近傍アミノ酸残基群抽出手段（近傍アミノ酸残基群抽出ステップ）と、上記近傍アミノ酸残基群の切り口の部分にキャップ用置換基を付加するキャップ付加手段（キャップ付加ステップ）と、上記キャップ付加手段（キャップ付加ステップ）により上記キャップ用置換基が付加された上記近傍アミノ酸残基群の全体の電荷を計算する電荷計算手段（電荷計算ステップ）と、上記キャップ付加手段（キャップ付加ステップ）により上記キャップ用置換基が付加された上記近傍アミノ酸残基群について、上記電荷計算手段（電荷計算ステップ）により計算された上記電荷を用いて上記特定のアミノ酸残基の原子座標について構造最適化を実行する構造最適化手段（構造最適化ステップ）と、上記構造最適化手段（構造最適化ステップ）にて最適化された上記原子座標を、上記タンパク質の座標データ上の対応する原子座標と置換する原子座標置換手段（原子座標置換ステップ）とを備えた（含む）ことを特徴とする。

- この装置、方法、および、プログラムによれば、タンパク質の座標データを取得し、タンパク質の座標データについて、特定のアミノ酸残基から所定の距離内に含まれる近傍アミノ酸残基群の座標を抽出し、近傍アミノ酸残基群の切り口の部分にキャップ用置換基を付加し、キャップ用置換基が付加された近傍アミノ酸残基群の全体の電荷を計算し、キャップ用置換基が付加された近傍ア

ミノ酸残基群について、計算された電荷を用いて特定のアミノ酸残基の原子座標について構造最適化を実行し、最適化された原子座標を、タンパク質の座標データ上の対応する原子座標と置換するので、水素位置の決定やパッキングの問題の解消を、実用的な計算資源を用いて行うことが可能となる。

- 5       また、この装置、方法、および、プログラムによれば、既存の計算プログラムには一切手を加えずに最適化処理の高速化を図ることができる。すなわち、既存の分子軌道計算プログラムや、分子力学計算プログラムの入出力ファイルを用いて、本装置を実行することができる。但し、本装置のアルゴリズムを既存の分子軌道計算プログラムや、分子力学計算プログラムに組み込むことも可能である。
- 10

また、この装置、方法、および、プログラムによれば、従来の方法では不可能な溶媒効果を考慮に入れたタンパク質の構造最適化が可能になる。

- つぎの発明にかかるタンパク質構造最適化装置、タンパク質構造最適化方法、および、プログラムは、上記に記載のタンパク質構造最適化装置、タンパク質構造最適化方法、および、プログラムにおいて、上記キャップ用置換基は、水素原子 (H) またはメチル基 ( $\text{CH}_3$ ) であることを特徴とする。
- 15

- これはキャップ用置換基の一例を一層具体的に示すものである。この装置、方法、および、プログラムによれば、キャップ用置換基は、水素原子 (H) またはメチル基 ( $\text{CH}_3$ ) であるので、近傍アミノ酸残基群について機械的に座標を切り取ったときの切り口が、ラジカルとなり計算に不都合を生じることを容易に解消することができるようになる。
- 20

- つぎの発明にかかるタンパク質構造最適化装置、タンパク質構造最適化方法、および、プログラムは、上記に記載のタンパク質構造最適化装置、タンパク質構造最適化方法、および、プログラムにおいて、上記近傍アミノ酸残基群抽出手段 (近傍アミノ酸残基群抽出ステップ) は、抽出した上記近傍アミノ酸残基群の中にシステイン (CYS) が含まれている場合には、当該システイン (CYS) とジスルフィド結合をしておりかつ上記近傍アミノ酸残基群には含まれ
- 25

ない別のシステイン（C Y S）が存在しているか判定し、当該別のシステイン（C Y S）が存在する場合には当該別のシステイン（C Y S）も近傍アミノ酸残基群に加えることを特徴とする。

5 これは近傍アミノ酸残基群抽出手段（近傍アミノ酸残基群抽出ステップ）の一例を一層具体的に示すものである。この装置、方法、および、プログラムによれば、近傍アミノ酸残基群抽出手段（近傍アミノ酸残基群抽出ステップ）は、抽出した近傍アミノ酸残基群の中にシステイン（C Y S）が含まれている場合には、当該システイン（C Y S）とジスルフィド結合をしておりかつ近傍アミノ酸残基群には含まれない別のシステイン（C Y S）が存在しているか判定し、  
10 当該別のシステイン（C Y S）が存在する場合には当該別のシステイン（C Y S）も近傍アミノ酸残基群に加えるので、システイン間のジスルフィド結合を考慮して構造最適化を行うことができるようになる。

また、本発明は記録媒体に関するものであり、本発明にかかる記録媒体は、上記に記載されたプログラムを記録したことを特徴とする。

15 この記録媒体によれば、当該記録媒体に記録されたプログラムをコンピュータに読み取らせて実行することによって、上記に記載されたプログラムをコンピュータを利用して実現することができ、これら各方法と同様の効果を得ることができる。

## 20 図面の簡単な説明

第1図は、本発明の基本原理を示す原理構成図であり、第2図は、本発明が適用される本システムの構成の一例を示すブロック図であり、第3図は、予測結果データベース106aに格納される情報の一例を示す図であり、第4図は、本実施形態における本システムのメイン処理の一例を示すフローチャートであり、  
25 第5図は、本実施形態における本システムの二次構造データ取得処理の一例を示すフローチャートであり、第6図は、本システムのフラストレーション計算部102eにより実行されるフラストレーション実行処理の一例を示すフ

ローチャートであり、第7図は、相互作用部位予測装置100の出力装置114に表示される相互作用部位予測結果の表示画面の一例を示す図であり、第8図は、相互作用部位予測装置100のモニタに表示される本実施例の処理結果出力画面の一例を示す図であり、第9図は、既知のドッキング・シミュレーションによりフラストレーションが大きい部位であると予測された部位が実際に相互作用部位として機能しているかを確認するための図であり、第10図は、本発明の基本原理を示す原理構成図であり、第11図は、本発明が適用される本システムの構成の一例を示すブロック図であり、第12図は、フロンティア軌道計算部1102aの構成の一例を示すブロック図であり、第13図は、活性部位予測部1102gの構成の一例を示すブロック図であり、第14図は、本実施形態における本システムのメイン処理の一例を示すフローチャートであり、第15図は、本実施形態における本システムの分子軌道計算処理の一例を示すフローチャートであり、第16図は、本実施形態における本システムのフロンティア軌道とその周辺軌道による候補アミノ酸残基決定処理の一例を示すフローチャートであり、第17図は、本実施形態における本システムの各分子軌道のアミノ酸への帰属情報決定処理の一例を示すフローチャートであり、第18図は、本実施形態における本システムの候補アミノ酸残基比較処理の一例を示すフローチャートであり、第19図は、本実施形態における本システムの主鎖の重原子に局在する軌道エネルギーによる候補アミノ酸残基決定処理の一例を示すフローチャートであり、第20図は、分子軌道計算で得られた計算結果の一例を示す図であり、第21図は、候補アミノ酸残基が蛋白質の立体構造中どの位置に存在しているかを確認するための表示画面の一例を示す図であり、第22図は、分子軌道計算で得られた計算結果の一例を示す図であり、第23図は、実施例1においてリボヌクレアーゼT1のフロンティア軌道が分布したアミノ酸残基を示す図表であり、第24図は、本実施例1において主鎖の窒素原子に分布している分子軌道の軌道エネルギーをアミノ酸の残基番号に対してプロットした図であり、第25図は、本実施例1において軌道エネルギーが高

いアミノ酸残基と軌道エネルギーとを抽出した図表であり、第26図は、本実施例1において第23図に示したフロンティア軌道による候補アミノ酸残基と、第24図および第25図に示した主鎖原子の軌道エネルギーによる候補アミノ酸残基の共通部分を抽出した図表であり、第27図は、実施例2においてリボヌクレアーゼAのフロンティア軌道が分布したアミノ酸残基を示す図表であり、第28図は、本実施例2において主鎖の窒素原子に分布している分子軌道の軌道エネルギーをアミノ酸の残基番号に対してプロットした図であり、第29図は、本実施例2において軌道エネルギーが高いアミノ酸残基と軌道エネルギーとを抽出した図表であり、第30図は、本実施例2において第27図に示したフロンティア軌道による候補アミノ酸残基と、第28図および第29図に示した主鎖原子の軌道エネルギーによる候補アミノ酸残基の共通部分を抽出した図表であり、第31図は、本発明の基本原理を示す原理構成図であり、第32図は、本発明が適用される本システムの構成の一例を示すブロック図であり、第33図は、本実施形態における本システムのメイン処理の一例を示すフローチャートであり、第34図は、本実施形態における本システムの溶媒接触面特定処理の一例を示すフローチャートであり、第35図は、本実施形態における本システムの疎水面特定処理の一例を示すフローチャートであり、第36図は、本実施形態における本システムの静電相互作用部位特定処理の一例を示すフローチャートであり、第37図は、本実施形態における本システムの相互作用部位特定処理の一例を示すフローチャートであり、第38図は、本実施形態における本システムの相互作用部位予測処理の一例を示すフローチャートであり、第39図は、タンパク質相互作用情報処理装置100が、溶媒接触面特定部102bの処理により、barnase-barstar複合体の結晶構造に基づき、barnaseについて各アミノ酸残基毎に溶媒接触面積の差 $\Delta S$ を計算した処理図であり、第40図は、タンパク質相互作用情報処理装置100が、疎水面特定部102cの処理により、barnase単体の結晶構造に基づき、barnaseについて各アミノ酸残基毎に疎水性相互作用エネルギーを計算

した処理図であり、第41図は、タンパク質相互作用情報処理装置100が、静電相互作用特定部102dの処理により、barnase単体の結晶構造に基づき、barnaseについて各アミノ酸残基毎に静電相互作用エネルギーを計算した処理図であり、第42図は、タンパク質相互作用情報処理装置100が、溶媒接触面特定部102bの処理により、barnase-barstar複合体の結晶構造に基づき、barstarについて各アミノ酸残基毎に溶媒接触面積の差 $\Delta S$ を計算した処理図であり、第43図は、タンパク質相互作用情報処理装置100が、疎水面特定部102cの処理により、barstar単体の結晶構造に基づき、barstarについて各アミノ酸残基毎に疎水性相互作用エネルギーを計算した処理図であり、第44図は、タンパク質相互作用情報処理装置100が、静電相互作用特定部102dの処理により、barstar単体の結晶構造に基づき、barstarについて各アミノ酸残基毎に静電相互作用エネルギーを計算した処理図であり、第45図は、タンパク質相互作用情報処理装置100が、溶媒接触面特定部102bの処理により、Ribonuclease-inhibitor複合体の結晶構造に基づき、Ribonucleaseについて各アミノ酸残基毎に溶媒接触面積の差 $\Delta S$ を計算した処理図であり、第46図は、タンパク質相互作用情報処理装置100が、疎水面特定部102cの処理により、Ribonuclease単体の結晶構造に基づき、Ribonucleaseについて各アミノ酸残基毎に疎水性相互作用エネルギーを計算した処理図であり、第47図は、タンパク質相互作用情報処理装置100が、静電相互作用特定部102dの処理により、Ribonuclease単体の結晶構造に基づき、Ribonucleaseについて各アミノ酸残基毎に静電相互作用エネルギーを計算した処理図であり、第48図は、タンパク質相互作用情報処理装置100が、溶媒接触面特定部102bの処理により、Ribonuclease-inhibitor複合体の結晶構造に基づき、inhibitorについて各アミノ酸残基毎に溶媒接触面積の差 $\Delta S$ を計算した処理図であり、第49図は、タンパク質相互作用情報

報処理装置100が、疎水面特定部102cの処理により、inhibitor単体の結晶構造に基づき、inhibitorについて各アミノ酸残基毎に疎水性相互作用エネルギーを計算した処理図であり、図50は、タンパク質相互作用情報処理装置100が、静電相互作用特定部102dの処理により、inhibitor単体の結晶構造に基づき、inhibitorについて各アミノ酸残基毎に静電相互作用エネルギーを計算した処理図であり、第51図は、本発明により1つの蛋白質のアミノ酸配列情報からその蛋白質の結合部位を予測する場合の概念を説明する図であり、第52図は、本発明により複数の蛋白質のアミノ酸配列情報からそれらの蛋白質で複合体を形成するときの結合部位を予測する場合の概念を説明する図であり、第53図は、本発明が適用される本システムの構成の一例を示すブロック図であり、第54図は、本発明が適用される空間距離決定部3102bの構成の一例を示すブロック図であり、第55図は、本発明が適用されるエネルギー計算部3102dの構成の一例を示すブロック図であり、第56図は、本発明の高速計算手法の概念を示す図であり、第57図は、複数のアミノ酸配列上で結合残基を仮定する場合の概念を示す図であり、第58図は、着目残基の概念を説明する図であり、第59図は、本実施形態における本システムの処理の一例を示すフローチャートであり、第60図は、処理結果の候補アミノ酸残基のエネルギーなどの一例を示す図であり、第61図は、立体構造上で不安定な部分がクラスター化している場合の一例を示す図であり、第62図は、ドッキングシミュレーションを用いて複合体構造を生成する場合の概念を示す図であり、第63図は、蛋白質Aと蛋白質Bの各アミノ酸残基を結合残基とした場合のエネルギーの総和をプロットした図の一例であり、第64図は、2つのグルタミン酸の配列上の距離と空間上の距離との関係を示す図であり、第65図は、リボヌクレアーゼAの各アミノ酸残基のエネルギーをアミノ酸残基番号に対してプロットした図であり、第66図は、リボヌクレアーゼAのアミノ酸残基のエネルギーが0以上のものを結合部位候補として表にまとめた図であり、第67図は、PDBに格納されたアセチルコ



リン・エステラーゼ・インヒビターの立体構造情報データの一部を示す図であり、第68図は、本発明により求めたアセチルコリン・エステラーゼ・インヒビターのエネルギーを示す図であり、第69図は、アセチルコリン・エステラーゼ・インヒビターのエネルギーが0以上のもの10個を結合部位候補として取り出し、実際に結合部位かどうかを実験的に調べた結果を示す図であり、第70図は、横軸に、`huntingtin-associated protein interacting protein`の結合部位のアミノ酸残基番号をとり、縦軸に`nitric oxide synthase 2A`の結合部位のアミノ酸残基番号をとり、各結合部位で複合体を生成したときのエネルギーの総和を等高線表示した図であり、第71図は、各候補蛋白質の相互作用エネルギーと遺伝子の数のヒストグラムであり、第72図は、本発明の基本原理を示すフローチャートであり、第73図は、本発明が適用される本システムの構成の一例を示すブロック図であり、第74図は、本実施形態における本システムのメイン処理の一例を示すフローチャートであり、第75図は、タンパク質の座標データの一例を示す図であり、第76図は、本実施形態における切り口に水素原子を付加する場合のキャップ付加処理の一例を示すフローチャートであり、第77図は、もとの座標とキャップ置換基を付加した後の座標の概念を示す図であり、第78図は、本実施形態における切り口に水素原子を付加する場合のキャップ付加処理の一例を示すフローチャートであり、第79図は、もとの座標とキャップ置換基を付加した後の座標の概念を示す図であり、第80図は、本実施形態における切り口にメチル基を付加する場合のキャップ付加処理の一例を示すフローチャートであり、第81図は、もとの座標とキャップ置換基を付加した後の座標の概念を示す図であり、第82図は、本実施形態における切り口にメチル基を付加する場合のキャップ付加処理の一例を示すフローチャートであり、第83図は、もとの座標とキャップ置換基を付加した後の座標の概念を示す図であり、第84図は、PDB形式データの三文字表記（18-20カラムの文字）でアミノ酸種別を判別する場合の概念を説明する

図であり、第 8 5 図は、アミノ酸残基 i の水素原子に最適化フラグを設定する場合の一例を示す図であり、第 8 6 図は、アミノ酸残基 i の水素と側鎖原子に最適化フラグを設定する場合の一例を示す図であり、第 8 7 図は、MOPAC 2 0 0 0 の入力ファイルの一例を示す図であり、第 8 8 図は、MOPAC 2 0 0 0 による構造最適化処理の結果を示す出力ファイルの一例を示す図であり、第 8 9 図は、従来の最適化手法（MOZYME 法＋BFGS 法）と本発明の手法により水素構造を最適化した場合の計算結果を示す図であり、第 9 0 図は、従来の最適化手法（MOZYME 法＋BFGS 法）と本発明の手法により側鎖構造を最適化した場合の計算結果を示す図である。

10

発明を実施するための最良の形態

（I）以下に、本発明にかかる相互作用部位予測装置、相互作用部位予測方法、プログラム、および、記録媒体の実施の形態を図面に基づいて詳細に説明する。なお、この実施の形態によりこの発明が限定されるものではない。

15 [本発明の概要]

以下、本発明の概要について説明し、その後、本発明の構成および処理等について詳細に説明する。図 1 は本発明の基本原理を示す原理構成図である。

本発明は、概略的に、以下の基本的特徴を有する。まず、利用者は、目的のタンパク質の一次配列情報である目的配列データ 1 0 を本発明の相互作用部位予測装置に入力する。この目的配列データ 1 0 の入力は、例えば、利用者が S W I S S - P R O T や P I R や T r E M B L 等の外部のデータベースに登録された一次配列情報を選択することにより入力してもよく、また、所望の一次配列情報を直接入力してもよい。

ついで、本発明の相互作用部位予測装置は、タンパク質の一次配列情報から当該タンパク質の二次構造を予測する二次構造予測プログラム 2 0 a ～ 2 0 d に対して入力された目的配列データ 1 0 の二次構造予測シミュレーションを実行する。ここで、二次構造プログラム 2 0 a ～ 2 0 d は、例えば、C h o u -

25

F a s m a nの方法や、n e u r a l n e t w o r kを使う手法や、l i n e a r s t a t i s t i c sを用いる手法や、n e a r e s t n e i g h b o r m e t h o dを用いる手法などを利用して二次構造予測シミュレーションを実行する。

- 5 ついで、本発明の相互作用部位予測装置は、各二次構造予測プログラム 2 0 a ~ 2 0 d の二次構造予測結果 3 0 a ~ 3 0 d を比較する ( 6 0 ) 。すなわち、目的配列データ 6 1 に対応する各予測プログラムの実行結果を並置して比較する ( 6 3 ~ 6 6 ) 。

そして、本発明の相互作用部位予測装置は、この比較結果に基づいて、目的  
10 のタンパク質の一次配列情報の局所部分のフラストレーションを計算する ( 7 0 ) 。すなわち、比較結果から各予測結果データ ( 6 3 ~ 6 6 ) で異なる二次構造を予測した局所部分を抽出して、その部分のフラストレーションを計算する。既存の二次構造予測プログラム 2 0 a ~ 2 0 d は、基本的には一次配列情報のうち一部の局所配列を見て予測を行うが、二次構造はタンパク質の全体の  
15 構造との関係で最終的に決定されるので、全体と局所部分との間に整合性が取れない部位、すなわちフラストレーションが大きい局所的部位では二次構造予測結果は外れる場合が多い。従って、複数のプログラムにおいて予測結果が外れる局所部分については、フラストレーションが大きいと仮定することができる。

- 20 フラストレーションの計算方法は、例えば、異なる予測結果データを出力した二次構造予測プログラム数に応じてフラストレーションを増減してもよく、あるいは、異なる予測結果のそれぞれの構造における確信度の平均値や分散値などに応じてフラストレーションを増減してもよく、また、その部分のアミノ酸配列を分子力学または分子動力学などの手法を用いてエネルギー量を求めそ  
25 のエネルギー量を用いてフラストレーションを計算してもよい。

そして、本発明の相互作用部位予測装置は、計算された局所部分のフラストレーションにより目的のタンパク質の相互作用部位を予測する ( 8 0 ) 。すな

わち、例えば、一定の閾値を超えるフラストレーションが存在する局所部分（67）を相互作用部位と予測する。

また、本発明の相互作用部位予測装置は、例えば、PDBやSCOPなどの外部のデータベースに目的のタンパク質の二次構造データが登録されている場合  
5 にはその二次構造データ40を取得し、予測結果の比較において用いる（60）。すなわち、目的のタンパク質が実際に採る二次構造データ62と、予測プログラムの予測結果データ63～66とを比較する。

そして、実際の二次構造データ62と予測プログラムの予測結果データ63～66とが異なっている部分については、高いフラストレーションを計算する。  
10 タンパク質の立体構造データが分かっている場合、すなわち、既存のPDBなどに立体構造データが登録されているタンパク質は、その全体構造が既知であるため、様々な二次構造予測手法の予測結果とその実際の構造との差を見ることにより、より明確にフラストレーションのある局所部位（相互作用部位になっている確率の高い部位）を見つけることが可能になる。例えば、実際の二次  
15 構造データ62と異なる予測結果データを出力した二次構造予測プログラム数に応じてフラストレーションを増減してもよい。

さらに、本発明の相互作用部位予測装置は、二次構造予測プログラム20a～20dの二次構造予測結果データ30a～30dに対する確信度を示す確信度情報50を設定する。すなわち、二次構造予測プログラム20a～20dの  
20 シミュレーション精度を実際の二次構造データなどに基づいて設定する。

そして、本発明の相互作用部位予測装置は、設定された確信度情報および比較結果に基づいて、局所部分の上記フラストレーションを計算する。すなわち、確信度情報が高い（すなわち、シミュレーションの精度の高い）プログラムによる二次構造予測結果データに対する重みを高くすることにより、フラストレーション計算においてシミュレーション結果に対する確信度を反映させること  
25 ができる。

[システム構成]

まず、本システムの構成について説明する。図2は、本発明が適用される本システムの構成の一例を示すブロック図であり、該構成のうち本発明に係る部分のみを概念的に示している。本システムは、概略的に、相互作用部位予測装置100と、配列情報や立体構造等に関する外部データベースおよびホモロジー検索や二次構造予測等の外部プログラム等を提供する外部システム200とを、ネットワーク300を介して通信可能に接続して構成されている。

図2においてネットワーク300は、相互作用部位予測装置100と外部システム200とを相互に接続する機能を有し、例えば、インターネット等である。

図2において外部システム200は、ネットワーク300を介して、相互作用部位予測装置100と相互に接続され、利用者に対して配列情報や立体構造等に関する外部データベース、および、ホモロジー検索やモチーフ検索や二次構造予測等の外部プログラムを実行するウェブサイトを提供する機能を有する。

ここで、外部システム200は、WEBサーバやASPサーバ等として構成してもよく、そのハードウェア構成は、一般に市販されるワークステーション、パーソナルコンピュータ等の情報処理装置およびその付属装置により構成してもよい。また、外部システム200の各機能は、外部システム200のハードウェア構成中のCPU、ディスク装置、メモリ装置、入力装置、出力装置、通信制御装置等およびそれらを制御するプログラム等により実現される。

図2において相互作用部位予測装置100は、概略的に、相互作用部位予測装置100の全体を統括的に制御するCPU等の制御部102、通信回線等に接続されるルータ等の通信装置（図示せず）に接続される通信制御インターフェース部104、入力装置112および出力装置114に接続される入出力制御インターフェース部108、および、各種のデータベースやテーブル（予測結果データベース106a～タンパク質構造データベース106c）を格納する記憶部106を備えて構成されており、これら各部は任意の通信路を介して通信可能に接続されている。さらに、この相互作用部位予測装置100は、ル

ータ等の通信装置および専用線等の有線または無線の通信回線を介して、ネットワーク 300 に通信可能に接続されている。

図 2 において記憶部 106 に格納される各種のデータベースやテーブル（予測結果データベース 106 a ～タンパク質構造データベース 106 c）は、固定ディスク装置等のストレージ手段であり、各種処理に用いる各種のプログラムやテーブルやファイルやデータベースやウェブページ用ファイル等を格納する。

これら記憶部 106 の各構成要素のうち、予測結果データベース 106 a は、二次構造予測プログラムの予測結果などに関する情報を格納する予測結果情報格納手段である。図 3 は、予測結果データベース 106 a に格納される情報の一例を示す図である。

この予測結果データベース 106 a に格納される情報は、図 3 に示すように、目的のタンパク質の一次配列情報（アミノ酸配列情報）である目的配列データ、タンパク質構造データベースから取得した目的配列データの二次構造データ、および、各二次構造予測プログラムの予測結果データを相互に関連付けて構成されている。

また、確信度情報データベース 106 b は、二次構造予測プログラムの二次構造予測結果データに対する確信度を示す確信度情報を格納する予測結果情報格納手段である。例えば、シミュレーション結果の精度の標準的な値（例えば、二次構造予測結果と実際の二次構造データとの一致率であるシミュレーション精度が 60 % の場合など）の確信度を 1 とし、標準値より精度が高い場合にはその精度に応じて確信度の値を大きくし、また、標準値より精度が低い場合にはその精度に応じて確信度の値を小さくしてもよい。さらに、二次構造プログラムごと、構造ごと、各配列におけるアミノ酸ごとに確信度を設定してもよい。すなわち、例えば、ある二次構造予測プログラムがある配列のあるアミノ酸についてその二次構造を予測するとき、その構造が  $\alpha$  構造である確信度、 $\beta$  構造である確信度等をそれぞれ別々に設定してもよい。

また、タンパク質構造データベース 106 c は、タンパク質の立体構造データを格納したデータベースである。タンパク質構造データベース 106 c は、インターネットを経由してアクセスする外部のタンパク質構造データベースであつてもよく、また、これらのデータベースをコピーしたり、オリジナルの配  
5 列情報を格納したり、さらに独自のアノテーション情報等を付加したりして作成したインハウスデータベースであつてもよい。

また、図 2 において、通信制御インターフェース部 104 は、相互作用部位予測装置 100 とネットワーク 300（またはルータ等の通信装置）との間における通信制御を行う。すなわち、通信制御インターフェース部 104 は、他  
10 の端末と通信回線を介してデータを通信する機能を有する。

また、図 2 において、入出力制御インターフェース部 108 は、入力装置 112 や出力装置 114 の制御を行う。ここで、出力装置 114 としては、モニタ（家庭用テレビを含む）の他、スピーカを用いることができる（なお、以下においては出力装置をモニタとして記載する）。また、入力装置 112 として  
15 は、キーボード、マウス、および、マイク等を用いることができる。また、モニタも、マウスと協働してポインティングデバイス機能を実現する。

また、図 2 において、制御部 102 は、OS（Operating System）等の制御プログラム、各種の処理手順等を規定したプログラム、および所要データを格納するための内部メモリを有し、これらのプログラム等により、種々の処理を実行するための情報処理を行う。制御部 102 は、機能概念的に、目的配列入力部 102 a、二次構造予測プログラム実行部 102 b、二次構造予測プログラム 102 c、予測結果比較部 102 d、フラストレーション計算部 102 e、相互作用部位予測部 102 f、二次構造データ取得部 102 g、および、確信度情報設定部 102 h を備えて構成されている。  
20

このうち、目的配列入力部 102 a は、目的のタンパク質の一次配列情報（目的配列データ）を入力する入力手段である。また、二次構造予測プログラム実行部 102 b は、二次構造予測プログラムに対して入力手段により入力さ  
25

れた一次配列情報（目的配列データ）の二次構造予測シミュレーションを実行させる二次構造予測プログラム実行手段である。また、二次構造予測プログラム 102 c は、タンパク質の一次配列情報から当該タンパク質の二次構造を予測する二次構造予測プログラムである。

- 5       また、予測結果比較部 102 d は、二次構造予測プログラムの二次構造予測結果を比較する予測結果比較手段、および、二次構造予測プログラムの二次構造予測結果と二次構造データ取得手段により取得した二次構造データとを比較する予測結果比較手段である。また、フラストレーション計算部 102 e は、
- 10       予測結果比較手段による比較結果に基づいて、目的のタンパク質の一次配列情報（目的配列データ）の局所部分のフラストレーションを計算するフラストレーション計算手段、および、確信度情報設定手段により設定された確信度情報および比較結果に基づいて、局所部分のフラストレーションを計算するフラストレーション計算手段である。

- また、相互作用部位予測部 102 f は、フラストレーション計算手段により
- 15       計算された局所部分のフラストレーションにより目的のタンパク質の相互作用部位を予測する相互作用部位予測手段である。また、二次構造データ取得部 102 g は、目的のタンパク質の二次構造データを取得する二次構造データ取得手段である。また、確信度情報設定部 102 h は、二次構造予測プログラムの二次構造予測結果に対する確信度を示す確信度情報を設定する確信度情報設定
- 20       手段である。なお、これら各部によって行なわれる処理の詳細については、後述する。

#### [システムの処理]

次に、このように構成された本実施の形態における本システムの処理の一例について、以下に図 4～図 7 を参照して詳細に説明する。

- 25       [メイン処理]

次に、メイン処理の詳細について図 4 を参照して説明する。図 4 は、本実施の形態における本システムのメイン処理の一例を示すフローチャートである。



まず、相互作用部位予測装置 100 は、目的配列入力部 102 a の処理により、利用者に対して目的のタンパク質の一次配列情報（目的配列データ）を入力させる（ステップ S A-1）。

5 ついで、相互作用部位予測装置 100 は、二次構造データ取得部 102 g の処理により、利用者が入力した目的配列データの二次構造データを取得する（ステップ S A-2）。

ここで、ステップ S A-2 において、二次構造データ取得部 102 g により実行される二次構造データ取得処理の詳細について図 5 を参照して説明する。図 5 は、本実施形態における本システムの二次構造データ取得処理の一例を示すフローチャートである。

まず、二次構造データ取得部 102 g は、タンパク質構造データベース 106 c を参照して目的配列データが登録されているか判定する（ステップ S B-1）。ステップ S B-1 において、目的配列データがタンパク質構造データベース 106 c に登録されている場合には、二次構造データ取得部 102 g は、  
15 タンパク質構造データベース 106 c から目的配列データの二次構造データを取得して予測結果データベース 106 a の所定の記憶領域に格納する（ステップ S B-2）。

一方、ステップ S B-1 において、目的配列データがタンパク質構造データベース 106 c に登録されていない場合には、二次構造データ取得部 102 g  
20 は、目的配列データに類似する配列のタンパク質の二次構造データがタンパク質構造データベース 106 c に存在するか判断する（ステップ S B-3）。すなわち、二次構造データ取得部 102 g は、例えば、配列間のホモロジーを判定するプログラムを用いて、目的配列データと、タンパク質構造データベース 106 c に登録された構造既知のタンパク質に対応する配列データとを比較し  
25 て、高い相同性を有する配列データ（目的配列データの一部に対応するものであってもよい）が存在するか否かを判断する。

ステップS B-3において、目的配列データに類似する配列のタンパク質の二次構造データがタンパク質構造データベース106cに存在する場合には、二次構造データ取得部102gは、類似部分の二次構造データを予測結果データベース106aの所定の記憶領域に格納する（ステップS B-4）。なお、  
5 目的配列データのうち一部分について二次構造データが存在する場合には、二次構造データが存在する部分について、当該二次構造データを予測結果データベース106aに格納する。

一方、ステップS B-3において、目的配列データに類似する配列のタンパク質の二次構造データがタンパク質構造データベース106cに存在しない場合  
10 には、二次構造データ取得処理が終了する。

再び図4に戻り、相互作用部位予測装置100は、二次構造予測プログラム実行部102bの処理により、目的配列データを1つまたは2つ以上の二次構造予測プログラム102cに実行させる（ステップS A-3）。すなわち、二次構造予測プログラム実行部102bは、例えば、目的配列データを所定のフォーマットに変換したり、また、目的配列データに所定のヘッダ情報などを追加したりすることにより、各二次構造予測プログラム102cの入力形式を合わせた後、二次構造予測プログラム102cを実行する。ここで、二次構造予測プログラム102cは、相互作用部位予測装置100の内部に存在するプログラムであってもよく、また、ネットワーク300を介してリモートで実行  
15 することができる外部システム200の外部プログラムであってもよい。

ついで、二次構造予測プログラム実行部102bは、各二次構造予測プログラム102cのシミュレーション結果である二次構造予測結果を予測結果データベース106aの所定の記憶領域に格納する（ステップS A-4）。

ついで、相互作用部位予測装置100は、予測結果比較部102dの処理により、予測結果データベース106aに格納された目的配列データに対する各  
25 二次構造予測プログラム102cの二次構造予測結果を比較する（ステップS A-5）。すなわち、予測結果比較部102dは、各二次構造予測プログラム

102cの二次構造予測結果について、目的配列データの先頭から最後まで  
各予測結果を比較する。なお、ステップSA-2において、二次構造予測プロ  
グラム実行部102bが目的配列データに対応する二次構造データを取得でき  
た場合、すなわち、予測結果データベース106aに目的配列データの二次構  
5 造データが格納されている場合には、二次構造データと各二次構造予測プロ  
グラム102cの二次構造予測結果とを比較する。

ついで、相互作用部位予測装置100は、フラストレーション計算部102  
eの処理により、目的配列データの局所部分のフラストレーションのスコアを  
計算する（ステップSA-6）。ここで、図6は、本システムのフラストレー  
10 ション計算部102eにより実行されるフラストレーション実行処理の一例を  
示すフローチャートである。

図6に示すように、フラストレーション計算部102eによるフラストレー  
ションのスコアの計算方法は、例えば、二次構造予測プログラムが異なる二次  
構造予測結果を出力した局所部分について、結果が異なる二次構造予測プロ  
15 グラムの数に応じてスコアを増減してもよく、あるいは、異なる予測結果のそれ  
ぞれの構造における確信度の平均値や分散値などに応じてフラストレーション  
を増減してもよく、また、二次構造予測プログラムが異なる二次構造予測結果  
を出力した局所部分について、アミノ酸配列を分子力学または分子動力学な  
20 の手法を用いてエネルギー量を求めそのエネルギー量を用いてフラストレーシ  
ョンを計算してもよい（ステップSC-1）。

また、フラストレーション計算部102eは、二次構造データと予測プロ  
グラムの二次構造予測結果とが異なっている部分について、高いフラストレー  
ションのスコアを計算してもよい（ステップSC-2）。例えば、二次構造デー  
タと異なる二次構造予測結果を出力した二次構造予測プログラムの数に応じて  
25 スコアを増減してもよい。

また、フラストレーション計算部102eは、確信度情報データベース10  
6bを参照して、予め確信度情報設定部102hの処理により格納された各二

次構造予測プログラム 102c の確信度情報を取得し、確信度情報に基づいてフラストレーションのスコアを計算してもよい（ステップ SC-3）。すなわち、フラストレーション計算部 102e は、シミュレーション精度の高い二次構造予測プログラム 102c の二次構造予測結果に高い重みを付けて、フラストレーションのスコアを計算する。

ここで、確信度情報設定部 102h による確信度情報の設定の一例を示す。まず、確信度情報設定部 102h は、各二次構造予測プログラム 102c の二次構造予測結果と、二次構造データとを比較して、各二次構造予測プログラム 102c の二次構造予測結果の精度（一致率）を計算する。そして、確信度情報設定部 102h は、各二次構造予測プログラム 102c の精度の平均値を標準の確信度情報（例えば、1）として設定し、平均値以上の精度については標準の確信度情報よりも高い値（例えば、1 より大きい数）を計算し、平均値以下の精度については標準の確信度情報よりも低い値（例えば、1 より小さい数）を計算して、確信度情報データベース 106b の所定の記憶領域に格納する。

なお、確信度情報設定部 102h は、各二次構造予測プログラム 102c の確信度情報を各配列中の各アミノ酸（残基）毎に設定してもよい。すなわち、各二次構造予測プログラム 102c による配列の予測結果につき配列中のアミノ酸毎に二次構造予測プログラム 102c の確信度情報を設定してもよい（例えば、配列中の一番目のアミノ酸について、プログラム A は、 $\alpha$  構造の確信度情報が 1.5 であり、 $\beta$  構造の確信度情報が 0.7 であり、その他の構造の確信度情報が 1.1 である等）。

また、確信度情報設定部 102h は、各二次構造予測プログラム 102c の確信度情報を構造（ $\alpha$  構造、 $\beta$  構造など）毎に設定してもよい。すなわち、各二次構造予測プログラム 102c によっては、特定の構造について高い精度を有するものや低い精度を有するものが存在するため、構造毎に二次構造予測プログラム 102c の確信度情報を設定してもよい（例えば、プログラム A は、

$\alpha$  構造の確信度情報が 1.5 であり、 $\beta$  構造の確信度情報が 0.7 であり、その他の構造の確信度情報が 1.1 である等)。

再び図 4 に戻り、相互作用部位予測装置 100 は、相互作用部位予測部 102 f の処理により、計算された局所部分のフラストレーションのスコアにより  
5 目的配列データのうち相互作用部位となる局所部分を予測する (ステップ SA-7)。すなわち、相互作用部位予測部 102 f は、例えば、一定の閾値を超えるフラストレーションのスコアである局所部分を相互作用部位と予測する。

ついで、相互作用部位予測装置 100 は、配列データの相互作用部位の予測結果を出力装置 114 に出力する (ステップ SA-8)。

10 ここで、図 7 は、相互作用部位予測装置 100 の出力装置 114 に表示される相互作用部位予測結果の表示画面の一例を示す図である。この図に示すように相互作用部位予測結果の表示画面は、例えば、目的配列データの配列情報の表示領域 MA-1、相互作用部位と予測される局所部分の表示領域 MA-2 および MA-3、相互領域部位と予測される局所部分のフラストレーションのスコアの表示領域 MA-4 および MA-5 等を含んで構成されている。これにて、  
15 メイン処理が終了する。

#### [実施例]

以下に本発明の実施例について、図 8 および図 9 を参照して詳細に説明する。

本実施例は、哺乳類アデニル酸シクラーゼ (Mammalian Adenylyl Cyclase) (PDB ID: 1CJK) (本明細書において  
20 「MAC」という。) のアミノ酸配列についてプログラム 1 および 2 を用いて二次構造予測を行い、その二次構造予測結果に基づいてフラストレーションの計算を行い、相互作用部位の予測を行った場合の一例である。

図 8 は、相互作用部位予測装置 100 のモニタに表示される本実施例の処理  
25 結果出力画面の一例を示す図である。この図に示すように処理結果出力画面は、例えば、MAC のアミノ酸配列が  $\beta$  ストランド構造である場合の確信度を示すグラフの表示領域 MB-1、MAC のアミノ酸配列が  $\alpha$  ヘリックス構造である

場合の確信度を示すグラフの表示領域MB-2、MACのアミノ酸配列がその他の二次構造である場合の確信度を示すグラフの表示領域MB-3、MACのアミノ酸配列の表示領域MB-4、フラストレーションの値が高いアミノ酸配列の断片領域（すなわち、相互作用部位である可能性が高い領域）を示す表示領域MB-5、プログラム1の二次構造予測結果の表示領域MB-6、プログラム2の二次構造予測結果の表示領域MB-7等を含んで構成されている。

本実施例において、フラストレーションの計算は、2つのプログラムが異なった二次構造予測を行い、その配列部分が比較的長く、かつ、それぞれの予測結果の確信度が比較的高いものについて、フラストレーションを大きく設定している。この他に、確信度を用いずに、直接二次構造の予測の差を用いてフラストレーションを計算することもできる。

図9は、既知のドッキング・シミュレーションによりフラストレーションが大きい部位であると予測された部位が実際に相互作用部位として機能しているかを確認するための図である。

図9においてMACについて予測した立体構造は空間充填模型（space fill）にて記載されている。なお、フラストレーションの大きい部位は濃い色で表示されている。また、図9においてMACとコンプレックスを作るその他のタンパク質は針金模型（wire frame）にて記載されている。図9に示すように、フラストレーションの大きい部位は、他のタンパク質との距離が比較的近く、これらの部位またはこれらと連続した配列の一部が相互作用部位である確率が高いことを示している。

#### [他の実施の形態]

さて、これまで本発明の実施の形態について説明したが、本発明は、上述した実施の形態以外にも、上記特許請求の範囲に記載した技術的思想の範囲内において種々の異なる実施の形態にて実施されてよいものである。

例えば、相互作用部位予測装置100がスタンドアローンの形態で相互作用部位予測を行う場合を一例に説明したが、相互作用部位予測装置100とは別

筐体で構成されるクライアント端末からの要求に応じて相互作用部位予測を行い、その予測結果を当該クライアント端末に返却するように構成してもよい。

また、実施形態において説明した各処理のうち、自動的に行なわれるものとして説明した処理の全部または一部を手動的に行うこともでき、あるいは、手動的に行なわれるものとして説明した処理の全部または一部を公知の方法で自動的に行うこともできる。

この他、上記文書中や図面中で示した処理手順、制御手順、具体的名称、各種の登録データや検索条件等のパラメータを含む情報、画面例、データベース構成については、特記する場合を除いて任意に変更することができる。

また、相互作用部位予測装置 100 に関して、図示の各構成要素は機能概念的なものであり、必ずしも物理的に図示の如く構成されていることを要しない。

例えば、相互作用部位予測装置 100 の各サーバが備える処理機能、特に制御部にて行なわれる各処理機能については、その全部または任意の一部を、CPU (Central Processing Unit) および当該 CPU にて解釈実行されるプログラムにて実現することができ、あるいは、ワイヤードロジックによるハードウェアとして実現することも可能である。なお、プログラムは、後述する記録媒体に記録されており、必要に応じて相互作用部位予測装置 100 に機械的に読み取られる。

また、このプログラムは、相互作用部位予測装置 100 に対して任意のネットワークを介して接続されたアプリケーションプログラムサーバに記録されてもよく、必要に応じてその全部または一部をダウンロードすることも可能である。

また、記憶部 106 に格納される各種のデータベース等（予測結果データベース 106 a 〜タンパク質構造データベース 106 c）は、RAM、ROM等のメモリ装置、ハードディスク等の固定ディスク装置、フレキシブルディスク、光ディスク等のストレージ手段であり、各種処理やウェブサイト提供に用いる

各種のプログラムやテーブルやファイルやデータベースやウェブページ用ファイル等を格納する。

また、相互作用部位予測装置 100 は、既知のパーソナルコンピュータ、ワークステーション等の情報処理端末等の情報処理装置にプリンタやモニタやイメージスキャナ等の周辺装置を接続し、該情報処理装置に本発明の方法を実現  
5 させるソフトウェア（プログラム、データ等を含む）を実装することにより実現してもよい。

さらに、相互作用部位予測装置 100 の分散・統合の具体的形態は図示のものに限られず、その全部または一部を、各種の負荷等に応じた任意の単位で、  
10 機能的または物理的に分散・統合して構成することができる。例えば、各データベースを独立したデータベース装置として独立に構成してもよく、また、処理の一部を CGI（Common Gateway Interface）を用いて実現してもよい。

また、本発明にかかるプログラムを、コンピュータ読み取り可能な記録媒体  
15 に格納することもできる。ここで、この「記録媒体」とは、フレキシブルディスク、光磁気ディスク、ROM、EPROM、EEPROM、CD-ROM、MO、DVD等の任意の「可搬用の物理媒体」や、各種コンピュータシステムに内蔵されるROM、RAM、HD等の任意の「固定用の物理媒体」、あるいは、LAN、WAN、インターネットに代表されるネットワークを介してプログラムを送信する場合の通信回線や搬送波のように、短期にプログラムを保持  
20 する「通信媒体」を含むものとする。

また、「プログラム」とは、任意の言語や記述方法にて記述されたデータ処理方法であり、ソースコードやバイナリコード等の形式を問わない。なお、

「プログラム」は必ずしも単一的に構成されるものに限られず、複数のモジュールやライブラリとして分散構成されるものや、OS（Operating  
25 System）に代表される別個のプログラムと協働してその機能を達成するものをも含む。なお、実施の形態に示した各装置において記録媒体を読み取る



ための具体的な構成、読み取り手順、あるいは、読み取り後のインストール手順等については、周知の構成や手順を用いることができる。

また、ネットワーク 300 は、相互作用部位予測装置 100 と外部システム 200 とを相互に接続する機能を有し、例えば、インターネットや、イントラネットや、LAN（有線／無線の双方を含む）や、VANや、パソコン通信網や、公衆電話網（アナログ／デジタルの双方を含む）や、専用回線網（アナログ／デジタルの双方を含む）や、CATV網や、IMT2000方式、GSM方式またはPDC／PDC-P方式等の携帯回線交換網／携帯パケット交換網や、無線呼出網や、Bluetooth等の局所無線網や、PHS網や、CS、BSまたはISDB等の衛星通信網等のうちいずれかを含んでもよい。すなわち、本システムは、有線・無線を問わず任意のネットワークを介して、各種データを送受信することができる。

以上詳細に説明したように、本発明によれば、目的のタンパク質の一次配列情報を入力し、タンパク質の一次配列情報から当該タンパク質の二次構造を予測する二次構造予測プログラムに対して入力された一次配列情報の二次構造予測シミュレーションを実行させ、二次構造予測プログラムの二次構造予測結果を比較し、比較結果に基づいて、目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算し、計算された局所部分のフラストレーションにより目的のタンパク質の相互作用部位を予測するので、タンパク質の一次配列情報においてフラストレーションのある局所部位を発見することにより相互作用部位を効果的に予測することができる相互作用部位予測装置、相互作用部位予測方法、プログラム、および、記録媒体を提供することができる。

また、本発明によれば、目的のタンパク質の一次配列情報を入力し、目的のタンパク質の二次構造データを取得し、タンパク質の一次配列情報から当該タンパク質の二次構造を予測する二次構造予測プログラムに対して入力された一次配列情報の二次構造予測シミュレーションを実行させ、二次構造予測プログラムの二次構造予測結果と、取得した二次構造データとを比較し、比較結果に

に基づいて、目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算し、計算された局所部分のフラストレーションにより目的のタンパク質の相互作用部位を予測するので、二次構造予測プログラムの予測結果と目的のタンパク質の実際の二次構造との差を見ることにより、より明確にフラストレーションのある局所部位（相互作用部位になっている確率の高い部位）を見つけることが可能になる相互作用部位予測装置、相互作用部位予測方法、プログラム、および、記録媒体を提供することができる。

さらに、本発明によれば、二次構造予測プログラムの二次構造予測結果に対する確信度を示す確信度情報を設定し、設定された確信度情報および比較結果に基づいて、局所部分のフラストレーションを計算するので、確信度情報が高い（すなわち、シミュレーションの精度の高い）プログラムによる二次構造予測結果データに対する重みを高くすることにより、フラストレーション計算においてシミュレーション結果に対する確信度を反映させることができる相互作用部位予測装置、相互作用部位予測方法、プログラム、および、記録媒体を提供することができる。

（Ⅱ）また、以下に、本発明にかかる活性部位予測装置、活性部位予測方法、プログラム、および、記録媒体の実施の形態を図面に基づいて詳細に説明する。なお、この実施の形態によりこの発明が限定されるものではない。本実施の形態は、蛋白質の活性部位予測に関する場合を一例に説明するが、当業者であれば本実施の形態の記載に基づいて本発明を生理活性ポリペプチドに容易に適用することができる。

#### [本発明の概要]

以下、本発明の概要について説明し、その後、本発明の構成および処理等について詳細に説明する。図10は本発明の基本原理を示す原理構成図である。

本発明は、概略的に、以下の基本的特徴を有する。すなわち、まず目的の蛋白質の立体構造データをPDB（Protein Data Bank）などの外部データベースなどから取得する（ステップS1）。

そして、蛋白質の立体構造データに基づいて分子軌道計算を行い、目的の蛋白質の立体構造データに基づいてフロンティア軌道（最高被占軌道（HOMO）または最低空軌道（LUMO））、および／または、主鎖原子の軌道エネルギーを求める（ステップS2）。

5       ここで、最高被占軌道（HOMO）または最低空軌道（LUMO）の軌道エネルギーは、例えば、市販のプログラムMOPAC2000（J. J. P. Stewart, Fujitsu Limited, Tokyo, Japan（1999））等を用いてAM1ハミルトニアン法等により計算することができる（ステップS21）。

10       また、分子軌道計算には半経験的分子軌道計算、非経験的分子軌道計算の他にも、密度汎関数計算を用いてもよい。現在の計算機の処理能力では半経験的分子軌道が好適であるが、将来的にはより精度の高い方法での応用が可能である。

ここで、発明者は計算条件を鋭意検討した結果、予測に必要な3つの計算条件を見つけることに成功した（ステップS3）。第一は、水分子を計算に含めることである。水分子と蛋白質との間の水素結合、さらには水分子と蛋白質との間の電荷移動を考慮するために、入力データの蛋白質のまわりに水分子を発生させることが必要である。結晶構造のデータのなかに水分子の情報が含まれているので、それらを活用することも可能であるが、ほとんどの場合は数が全く足りない。従って、例えば蛋白質と水素結合できるような位置に水分子を置くという方法等により入力データの蛋白質のまわりに水分子を発生させて分子軌道計算を行う（ステップS31）。

第二に、水分子の誘電的な効果を考慮する必要がある（ステップS32）。これには様々な方法が考えられる。例えば、蛋白質の周りに連続的な誘電体を置く方法（Klamtらが開発したCOSMO法などが代表例である。）等を用いることができる。

第三に、非常に大きな分子について本発明を応用したい場合は、溶媒の影響を考慮することにより計算機の処理能力の限界を超えてしまうことが予想される。その場合は、蛋白質表面の解離性アミノ酸残基を無電荷の状態（例えば、グルタミン酸をプロトン化する）にし、内部に埋まっている解離性アミノ酸を電荷状態（例えば、グルタミン酸を脱プロトン化する）にすることによって、溶媒を考慮した計算結果を近似的に求める（ステップS 33）。

このように、本発明では、3つの計算条件を適宜設定することにより、効率的に分子軌道計算を実行することができ、さらに、活性部位予測の精度を大幅に向上させることができるようになる。

ここで、本発明における「フロンティア軌道の周辺軌道」の定義を次のように行う。一般的に『フロンティア軌道』とは『最高被占軌道（HOMO）』と『最低空軌道（LUMO）』の2つのことを指す。しかし、蛋白質のような巨大分子の系ではフロンティア軌道とエネルギー的にほとんど変わらない分子軌道がフロンティア軌道と同様に機能に重要な役割を果たすことが多い。発明者が鋭意研究した結果、多少のエネルギーの違い（例えば1～2 eV程度）であれば、フロンティア軌道と変わらない役割を果たすことが分かった。そこで、本発明ではフロンティア軌道をその周辺領域にまで拡張している。例えば、最高被占軌道（HOMO）とのエネルギー差が予め定めた閾値（例えば、2 eV等）以内の全ての被占軌道、および、最低空軌道とのエネルギー差が予め定めた閾値（例えば、2 eV等）以内の全ての空軌道をフロンティア軌道の『周辺軌道』と定義する。この定義の拡張が本発明の特徴のうちの一つである。

次に、本発明は、求めたフロンティア軌道と周辺軌道を蛋白質のアミノ酸配列中の特定のアミノ酸残基に帰属させる（ステップS 4）。分子軌道のアミノ酸残基への帰属は以下のように行う。

各分子軌道は以下のように基底関数の線形結合によって表される。

$$\phi = \sum c_i \Phi_i$$

( $i$  は基底関数の番号、 $\Phi_i$  は基底関数、 $c_i$  は係数)

各基底関数は原子に所属し、各原子はアミノ酸残基に所属している。したがって、各基底関数はいずれかのアミノ酸残基に所属している。そこで、原子ごと、ならびにアミノ酸残基ごとの分布率を求める。

$$D(K) = \sum c_i^2$$

( $i$  は原子またはアミノ酸残基 $K$ に所属している全ての基底関数)

これにより、分子軌道ごとに最も分布率の大きなアミノ酸残基、または、最も分布率の大きな原子を含んでいるアミノ酸残基を得ることができる。それらを各分子軌道が分布するアミノ酸残基と定義する。このように定義すると、分子軌道がどこのアミノ酸に分布しているかを1対1で対応をとることができる。一般的に分子軌道はある程度の広がりがあるので、一つのアミノ酸残基に分布しているという発想は量子化学の世界では一般的ではないが、発明者は機能に関係する軌道に限っては、ほとんど一つのアミノ酸に局在しているという事実を発見した。分子軌道とアミノ酸を1対1で対応させることは、専門家以外にも分かりやすく、専門家以外が本発明を利用することには大きな助けとなる。この点も本発明の利点の一つである。

以上により、蛋白質のフロンティア軌道と周辺軌道が分布しているアミノ酸残基が判明するが、本発明は、このアミノ酸残基を活性部位の候補のアミノ酸残基（以下「候補アミノ酸残基」または単に「候補」という。）として決定する（ステップS4）。

次に、本発明は、活性部位になりえない候補の削除などを行い、活性部位を予測する（ステップS5）。例えば、トリプトファンやフェニルアラニンのような芳香環を含んでいるアミノ酸残基は性質上フロンティア軌道と周辺軌道になりやすいという性質をもっている。しかし、ほとんどの場合はそれらが活性

部位にはなりえないことがわかっている。同様にジスルフィド結合しているシステイン、メチオニンもフロンティア軌道と周辺軌道が分布しやすいが活性部位にはほとんどならないことがわかっている。フロンティア軌道と周辺軌道のうち、これらのアミノ酸残基に所属しているものは、活性部位の候補から除く。

- 5 残ったフロンティア軌道と周辺軌道が分布しているアミノ酸残基が活性部位の候補であるが、ここで、活性部位は一つのアミノ酸残基であるケースはほとんどなく、複数のアミノ酸残基によって構成されている。したがって、実際に目的の蛋白質の立体構造データから立体構造を既知のグラフィックソフトなどを用いて表示させて、フロンティア軌道と周辺軌道を表示させると多くの場合
- 10 で、フロンティア軌道と周辺軌道が集まって存在している場所がある。そのように、立体構造上で局在化してクラスタを作っている部分の候補アミノ酸残基が活性部位である可能性が極めて高いため、このような候補を選択して、活性部位として予想する。

- また、主鎖原子の軌道エネルギーも用いる場合には、上述したフロンティア
- 15 軌道を用いる場合と同様の計算条件で計算をするが、分子軌道をアミノ酸ではなく、分子に帰属させることに違いがある（ステップS 2 2）。すなわち、あるアミノ酸の主鎖の原子（例えば、窒素や炭素など）に分布している分子軌道の軌道エネルギーに着目する。そのような分子軌道は複数あるので、その中で最も特徴的な、例えば、最もエネルギーの高い被占軌道の軌道エネルギーに着
- 20 目する。ここでも、アミノ酸と軌道エネルギーが1対1で対応する。

- このように各アミノ酸を主鎖の原子に分布している分子軌道の軌道エネルギーと各アミノ酸を対応させて解析に応用するという方法は従来にはない手法である。例えば、アミノ酸の番号と軌道エネルギーをプロットとすると、軌道エネルギーの相対的な大きさが分かる。軌道エネルギーが相対的に高い原子が存在するアミノ酸残基の部分が活性部位である可能性が高い。また、予め定めた
- 25 値を超える軌道エネルギーをもつ分子軌道の分布するアミノ酸残基が活性部位

になる可能性が高い。閾値は同様の機能を持つ蛋白質の活性部位の軌道エネルギーなどを参考にして決定してもよい。

以上を示した2つの手法（ステップS 2 1およびステップS 2 2）は、それぞれ活性部位を予測すること、分子軌道計算を利用していることで共通している。しかしながら、その予測方法による予測結果は全く同一ではない。それぞれの方法に関して、強みや弱みがあることは容易に予想できる。これらを手法を組み合わせることで、それぞれの候補を比較することで、さらに精度を上げることができる。例えば、異なる方法の全ての予測結果において活性部位と予測されるアミノ酸残基、一部の方法で活性部位であると予測されるアミノ酸残基というように分類することで、活性部位らしさをより正確に表現することができる。

#### [システム構成]

まず、本システムの構成について図1 1～図1 3を参照して説明する。図1 1は、本発明が適用される本システムの構成の一例を示すブロック図であり、該構成のうち本発明に関係する部分のみを概念的に示している。本システムは、概略的に、蛋白質活性部位予測装置1 1 0 0と、蛋白質の構造情報等に関する外部データベースやホモロジー検索等の外部プログラム等を提供する外部システム1 2 0 0とを、ネットワーク1 3 0 0を介して通信可能に接続して構成されている。

図1 1においてネットワーク1 3 0 0は、蛋白質活性部位予測装置1 1 0 0と外部システム1 2 0 0とを相互に接続する機能を有し、例えば、インターネット等である。

図1 1において外部システム1 2 0 0は、ネットワーク1 3 0 0を介して、蛋白質活性部位予測装置1 1 0 0と相互に接続され、利用者に対して蛋白質の構造情報等に関する外部データベースやホモロジー検索やモチーフ検索等の外部プログラムを実行するウェブサイトを提供する機能を有する。

ここで、外部システム1 2 0 0は、WEBサーバやASPサーバ等として構成してもよく、そのハードウェア構成は、一般に市販されるワークステーション

ン、パーソナルコンピュータ等の情報処理装置およびその付属装置により構成してもよい。また、外部システム 1 2 0 0 の各機能は、外部システム 1 2 0 0 のハードウェア構成中の CPU、ディスク装置、メモリ装置、入力装置、出力装置、通信制御装置等およびそれらを制御するプログラム等により実現される。

5 図 1 1 において蛋白質活性部位予測装置 1 1 0 0 は、概略的に、蛋白質活性部位予測装置 1 1 0 0 の全体を統括的に制御する CPU 等の制御部 1 1 0 2、通信回線等に接続されるルータ等の通信装置（図示せず）に接続される通信制御インターフェース部 1 1 0 4、入力装置 1 1 1 2 や出力装置 1 1 1 4 に接続される入出力制御インターフェース部 1 1 0 8、および、各種のデータベース  
10 やテーブルなどを格納する記憶部 1 1 0 6 を備えて構成されており、これら各部位は任意の通信路を介して通信可能に接続されている。さらに、この蛋白質活性部位予測装置 1 1 0 0 は、ルータ等の通信装置および専用線等の有線または無線の通信回線を介して、ネットワーク 1 3 0 0 に通信可能に接続されている。

記憶部 1 1 0 6 に格納される各種のデータベースやテーブル（蛋白質構造データベース 1 1 0 6 a および処理結果データ 1 1 0 6 b）は、固定ディスク装置等のストレージ手段であり、各種処理に用いる各種のプログラムやテーブルやファイルやデータベースやウェブページ用ファイル等を格納する。

これら記憶部 1 1 0 6 の各構成要素のうち、蛋白質構造データベース 1 1 0 6 a は、蛋白質の構造データ（アミノ酸配列データ、3 次元構造データ、各種  
20 のアノテーション情報等を含む）を格納したデータベースである。蛋白質構造データベース 1 1 0 6 a は、インターネットを経由してアクセスする外部のデータベースであってもよく、また、これらのデータベースをコピーしたり、オリジナルの配列情報を格納したり、さらに独自のアノテーション情報等を付加したりして作成したインハウスデータベースであってもよい。

25 また、処理結果データ 1 1 0 6 b は、制御部 1 1 0 2 による処理結果に関する情報等を格納する処理結果データ格納手段である。



また、図 11 において、通信制御インターフェース部 1104 は、蛋白質活性部位予測装置 1100 とネットワーク 1300（またはルータ等の通信装置）との間における通信制御を行う。すなわち、通信制御インターフェース部 1104 は、他の端末と通信回線を介してデータを通信する機能を有する。

5      また、図 11 において、入出力制御インターフェース部 1108 は、入力装置 1112 や出力装置 1114 の制御を行う。ここで、出力装置 1114 としては、モニタ（家庭用テレビを含む）の他、スピーカを用いることができる（なお、以下においては出力装置 1114 をモニタとして記載する場合がある）。また、入力装置 1112 としては、キーボード、マウス、および、マイク等を用いることができる。また、モニタも、マウスと協働してポインティングデバイス機能を実現する。

15      また、図 11 において、制御部 1102 は、OS (Operating System) 等の制御プログラム、各種の処理手順等を規定したプログラム、および所要データを格納するための内部メモリを有し、これらのプログラム等により、種々の処理を実行するための情報処理を行う。制御部 1102 は、機能概念的に、フロンティア軌道計算部 1102a、周辺軌道決定部 1102b、水分子設定部 1102c、誘電体設定部 1102d、電荷設定部 1102e、候補アミノ酸残基決定部 1102f、活性部位予測部 1102g、軌道エネルギー計算部 1102h、および、構造データ取得部 1102p を備えて構成さ

20      れている。

このうち、フロンティア軌道計算部 1102a は、構造データに基づいて蛋白質の電子状態を分子軌道計算によって求めてフロンティア軌道を特定するフロンティア軌道計算手段である。ここで、フロンティア軌道計算部 1102a は、図 12 に示すように、最高被占軌道計算部 1102i と最低空軌道計算部

25      1102j を備えて構成される。

また、周辺軌道決定部 1 1 0 2 b は、フロンティア軌道と予め定めたエネルギー差になる分子軌道を当該フロンティア軌道の周辺軌道として決定する周辺軌道決定手段である。

また、水分子設定部 1 1 0 2 c は、蛋白質のまわりに水分子を発生させて分子軌道計算等の量子化学計算を行う水分子設定手段である。

また、誘電体設定部 1 1 0 2 d は、蛋白質の周りに連続的な誘電体を置いて分子軌道計算等の量子化学計算を行う誘電体設定手段である。

また、電荷設定部 1 1 0 2 e は、蛋白質表面の解離性アミノ酸残基を無電荷の状態にし、内部に埋まっている解離性アミノ酸を電荷状態にして分子軌道計算等の量子化学計算を行う電荷設定手段である。

また、候補アミノ酸残基決定部 1 1 0 2 f は、フロンティア軌道と周辺軌道が分布しているアミノ酸残基、および／または、軌道エネルギーのうち予め定めた値を超える軌道エネルギーとなる分子軌道および／または相対的に軌道エネルギーが高い分子軌道が分布しているアミノ酸残基を活性部位の候補のアミノ酸残基として決定する候補アミノ酸残基決定手段である。

また、活性部位予測部 1 1 0 2 g は、候補アミノ酸残基決定部 1 1 0 2 f により決定された候補アミノ酸残基の中から活性部位を選択して活性部位を予測する活性部位予測手段である。ここで、活性部位予測部 1 1 0 2 g は、図 1 3 に示すように、活性部位になりえない候補の削除を行う特定アミノ酸残基除外部 1 1 0 2 k と、立体構造上で局在化してクラスタを作っている部分の候補アミノ酸残基を選択する局在アミノ酸残基選択部 1 1 0 2 m と、各手法による候補を比較して重複する候補を選択する候補比較部 1 1 0 2 n とを備えて構成される。

また、構造データ取得部 1 1 0 2 p は、目的の蛋白質の構造データを取得する構造データ取得手段である。

なお、これら各部によって行なわれる処理の詳細については、後述する。

[システムの処理]

次に、このように構成された本実施の形態における本システムの処理の一例について、以下に図 1 4～図 2 1 を参照して詳細に説明する。

[メイン処理]

まず、メイン処理の詳細について図 1 4 を参照して説明する。図 1 4 は、本  
5 実施形態における本システムのメイン処理の一例を示すフローチャートである。

蛋白質活性部位予測装置 1 1 0 0 は、構造データ取得部 1 1 0 2 p の処理により、まず目的の蛋白質の立体構造データを P D B ( P r o t e i n D a t a B a n k ) などの外部データベースなどから取得する (ステップ S A 1 - 1 ) 。

10 ついで、蛋白質活性部位予測装置 1 1 0 0 は、制御部 1 1 0 2 の処理により、蛋白質の立体構造データに基づいて量子化学計算による分子軌道の計算を行う (ステップ S A 1 - 2 ) 。ここで、分子軌道計算処理の詳細について図 1 5 を参照して説明する。図 1 5 は、本実施形態における本システムの分子軌道計算処理の一例を示すフローチャートである。

15 まず、蛋白質活性部位予測装置 1 1 0 0 は、蛋白質の座標を取得し (ステップ S B 1 - 1 ) 、分子軌道計算を行う。ここで、分子軌道計算については、例えば、「計算機化学入門」 ( 櫻井実、猪飼篤 編、丸善、1 9 9 9 年 ) など詳しく解説されている。以下に分子軌道計算処理の一例を説明する。まず、フォック方程式を解く (ステップ S B 1 - 2 ～ステップ S B 1 - 7 ) 。この方程式は『非線形』であるため、解が収束するまで繰り返し計算する解法を取る。  
20

$$F C = S C \varepsilon$$

本方程式において、F はフォック行列、C は L C A O 係数を要素とした行列、  
25 S は重なり積分を要素とした行列、 $\varepsilon$  は軌道エネルギーの要素としたベクトルを示す。フォック行列は  $F = h + G \cdot D$  というように、密度行列 D と関連付けることができる。密度行列は L C A O 係数から計算することができる。F の生

成（ステップS B 1－4）、Fの対角化（ステップS B 1－5）、密度行列の生成（ステップS B 1－6）の各ステップを密度行列が収束するまで行う。

そして、蛋白質活性部位予測装置 1 1 0 0 は、軌道エネルギーや分子軌道の係数を取得し（ステップS B 1－8）、系のエネルギーを求める（ステップS  
5 B 1－9）。これにて、分子軌道計算処理が終了する。

再び図 1 4 に戻り、蛋白質活性部位予測装置 1 1 0 0 は、ステップS A 1－2において求めた分子軌道等の情報に基づいてフロンティア軌道とその周辺軌道から候補アミノ酸残基を決定する（ステップS A 1－3）。ここで、フロンティア軌道とその周辺軌道による候補アミノ酸残基決定処理の詳細について図  
10 1 6 を参照して説明する。図 1 6 は、本実施形態における本システムのフロンティア軌道とその周辺軌道による候補アミノ酸残基決定処理の一例を示すフローチャートである。

まず、蛋白質活性部位予測装置 1 1 0 0 は、計算した分子軌道が蛋白質のアミノ酸配列のうちどのアミノ酸残基に分布しているかを帰属する（ステップS  
15 C 1－1）。ここで、分子軌道計算を行うと、出力として、各分子軌道について2つの情報『分布の仕方』、『軌道エネルギー』が得られるが、ここでは『分布の仕方』の情報より、各分子軌道がどの原子（アミノ酸残基）に分布しているかを特定する。ここで、各分子軌道のアミノ酸への帰属情報決定処理の詳細について図 1 7 を参照して説明する。図 1 7 は、本実施形態における本シ  
20 ステムの各分子軌道のアミノ酸への帰属情報決定処理の一例を示すフローチャートである。

まず、N番目の分子軌道を取得し（ステップS D 1－1）、原子ごとに原子に所属している基底関数の係数の2乗を足し合わせ（ステップS D 1－2）、アミノ酸ごとにアミノ酸に所属する原子に所属する基底関数の係数の2乗和を  
25 足し合わせる（ステップS D 1－3）。

そして、和が最も大きいアミノ酸をN番目の分子軌道が所属しているアミノ酸であると特定する（ステップS D 1－4）。

また、図 20 は分子軌道計算で得られた計算結果の一例を示す図である。図 20 に示す例においては、5 残基からなるオリゴペプチド (R E W T Y) を一例に説明する。本図によると、分子軌道 1 はアミノ酸残基 R、分子軌道 2 はアミノ酸残基 T、分子軌道 3 はアミノ酸残基 E、分子軌道 4 はアミノ酸残基 W、分子軌道 5 はアミノ酸残基 R、分子軌道 6 はアミノ酸残基 Y、分子軌道 7 はアミノ酸残基 E に帰属する。

これにて、各分子軌道のアミノ酸への帰属情報決定処理が終了する。

再び図 16 に戻り、蛋白質活性部位予測装置 1100 は、フロンティア軌道とその周辺軌道を定義する。すなわち、フロンティア軌道計算部 1102a は、最高被占軌道計算部 1102i および最低空軌道計算部 1102j の処理により最高被占軌道 (HOMO) は分子軌道 4 に決定し、最低空軌道 (LUMO) は分子軌道 5 に決定する。また、本実施形態において 2 eV 以内の軌道エネルギーである分子軌道をフロンティア軌道の周辺軌道と定義すると、周辺軌道決定部 1102b は、分子軌道 2、3、4、5、6 までを周辺軌道と決定する。従って、候補アミノ酸残基決定部 1102f は、分子軌道 2、3、4、5、6 に対応するアミノ酸残基が、活性部位の候補アミノ酸残基として決定する (ステップ SC1-2)。

次に、活性部位予測部 1102g は、特定アミノ酸残基除外部 1102k の処理により、機能部位候補として不適切なものを除外する (ステップ SC1-3)。この例では、特定アミノ酸残基除外部 1102k は、分子軌道 4 は活性部位になる可能性が低いアミノ酸残基であるトリプトファンに分布しているので除外する。その結果、候補アミノ酸残基は分子軌道 2、3、5 および 6 に絞られる。

次に、活性部位予測部 1102g は、局在アミノ酸残基選択部 1102m の処理により、候補が空間的にどのように存在しているかを調べ、局在しているアミノ酸残基を選択する (ステップ SC1-4)。図 21 は、候補アミノ酸残

基が蛋白質の立体構造中どの位置に存在しているかを確認するための表示画面の一例を示す図である。

図 2 1 に示すように、既知のグラフィック表示プログラムにより、蛋白質の構造データを針金モデル、リボンモデル、パイプモデル、ボールアンドスティックモデル、または、空間充填モデルのうちいずれかのモデルによりグラフィック表示し、その中に候補アミノ酸残基を表示する。本図の場合は右のほうに偏ったクラスタリングが存在しているので、クラスタリングを形成する 3 つの候補が機能部位である可能性は高い。

これにて、フロンティア軌道とその周辺軌道による候補アミノ酸残基決定処理が終了する。

再び図 1 4 に戻り、蛋白質活性部位予測装置 1 1 0 0 は、ステップ S A 1 - 2 において求めた分子軌道等の情報に基づいて主鎖の重原子に局在する軌道エネルギーから候補アミノ酸残基を決定する（ステップ S A 1 - 4）。ここで、主鎖の重原子に局在する軌道エネルギーによる候補アミノ酸残基決定処理の詳細について図 1 9 を参照して説明する。図 1 9 は、本実施形態における本システムの主鎖の重原子に局在する軌道エネルギーによる候補アミノ酸残基決定処理の一例を示すフローチャートである。

まず、蛋白質活性部位予測装置 1 1 0 0 は、計算した分子軌道が蛋白質のアミノ酸配列を構成する原子のうちどの原子に分布しているかを帰属する（ステップ S F 1 - 1）。ステップ S C 1 - 1 においてはアミノ酸ごとの分布を求めたが、ここでは原子ごとの分布を求める点が異なる。

ここで、図 2 2 は分子軌道計算で得られた計算結果の一例を示す図である。本図によると、分子軌道 1 は原子番号 1、分子軌道 2 は原子番号 4、分子軌道 5 は原子番号 1、分子軌道 6 は原子番号 4、分子軌道 7 は原子番号 2、分子軌道 8 は原子番号 3、分子軌道 9 は原子番号 1、分子軌道 1 0 は原子番号 4 に帰属する。

次に、軌道エネルギー計算部 1102h は、特定の主鎖の重原子に帰属している分子軌道のみを抽出する（ステップ S F 1-2）。図 22 に示す例において、主鎖 N 原子に着目すると、R の主鎖 N 原子（原子番号 1）には、分子軌道 1、5、9 が分布し、E の主鎖 N 原子（原子番号 4）には、分子軌道 2、6、10 が分布する。

次に、軌道エネルギー計算部 1102h は、着目したものの中で、最もエネルギーの高い被占軌道を選択する（ステップ S F 1-3）。図 22 に示す例において、分子軌道 9 および 10 は空軌道なので除外すると、R の主鎖 N 原子（原子番号 1）は分子軌道 5、E の主鎖 N 原子（原子番号 4）は分子軌道 6 が最もエネルギーが高いため軌道エネルギー計算部 1102h はそれぞれ選択する。すなわち、R の軌道エネルギーについては -6 e V、E の軌道エネルギーについては -5 e V が代表エネルギーになる。

次に、軌道エネルギー計算部 1102h は、アミノ酸残基番号を横軸、代表エネルギーを縦軸にとり、代表エネルギーをプロットするプロット図を作成して（ステップ S F 1-4）、図中のピーク部分の周辺部位を候補アミノ酸残基として特定する（ステップ S F 1-5）。

これにて、主鎖の重原子に局在する軌道エネルギーによる候補アミノ酸残基決定処理が終了する。

再び図 14 に戻り、蛋白質活性部位予測装置 1100 は、活性部位予測部 1102g の処理により、候補アミノ酸残基の中から活性部位を選択して活性部位を予測する（ステップ S A 1-5）。ここで、候補アミノ酸残基比較処理の詳細について図 18 を参照して説明する。図 18 は、本実施形態における本システムの候補アミノ酸残基比較処理の一例を示すフローチャートである。

図 18 に示すように、上述したフロンティア軌道および主鎖原子の軌道エネルギーを用いた手法により複数の候補アミノ酸残基を生成すると（ステップ S E 1-1）、活性部位予測部 1102g は、候補比較部 1102n の処理により、各手法による候補が一致するか判定し（ステップ S E 1-2）、一致しな

い場合には、前後のアミノ酸も候補に追加（一致しない場合には更に次のアミノ酸も追加）して上述の手法による候補決定手法を再び実行する（ステップS E 1－3）。

一方、ステップS E 1－2において各手法による候補が一致する場合には、  
5 活性部位予測部1102gはこの候補を活性部位として予測する（ステップS E 1－4）。これにて、候補アミノ酸残基比較処理が終了する。

これにて、メイン処理が終了する。

[本発明の実施例1；リボヌクレアーゼT1]

次に、本発明の実施例1の詳細について図23～図26を参照して説明する。

10 加水分解酵素であるリボヌクレアーゼT1は実験的によく調べられていて、必須なアミノ酸残基はH i s 4 0、G l u 5 8、A r g 7 7、H i s 9 2であるということが実験的に証明されている。

リボヌクレアーゼT1について、X線結晶構造データに基づき市販のプログラムI n s i g h t I Iによって水素分子を付加して、分子軌道計算に必要な座  
15 標を完成した。そして、市販のプログラムM O P A C 2 0 0 0を用いて最適化構造を求めた後に、電子状態を得た。蛋白質まわりに水分子を置き、さらに連続誘電体近似（C O S M O法）を用いて溶媒の影響を考慮した。

ここで、実施例1においてリボヌクレアーゼT1のフロンティア軌道が分布したアミノ酸残基を図23に示す表に示す。

20 図23に示すように、活性部位の候補となりうるのは、HOMOから2番目にG l u 5 8と、HOMOから3番目にH i s 4 0、LUMOから4番目のH i s 9 2、LUMOから3番目にA r g 7 7が分布した。この4つのアミノ残基は近くに集まっているので、ここが活性部位であると容易に予想することができる。これは実験データと良く一致している。また、H i s 4 0とG l u 5  
25 8が求核的に、A r g 7 7とH i s 9 2が求電子的に働くと予想できる。つまり、従来の技術とは異なり、活性部位の位置だけではなく、反応のメカニズムを解析することができる。



次に主鎖の窒素原子に着目した。図 2 4 は、本実施例 1 において主鎖の窒素原子に分布している分子軌道の軌道エネルギーをアミノ酸の残基番号に対してプロットした図である。本図に示すようにアミノ酸残基番号 4 0、6 0、8 0、9 0 付近に軌道エネルギーの高い部分が現れる。また、本実施例 1 において軌道エネルギーが高いアミノ酸残基と軌道エネルギーとを抽出した表を図 2 5 に示す。軌道エネルギーの高いアミノ酸残基の周辺にあるアミノ酸残基が活性部位の候補にある。

また、図 2 3 に示したフロンティア軌道による候補アミノ酸残基と、図 2 4 および図 2 5 に示した主鎖原子の軌道エネルギーによる候補アミノ酸残基の共通部分を抽出した表を図 2 6 に示す。例えば、フロンティア軌道による方法で求核基の候補 4 つ、求電子基の候補 4 つを挙げた。また、主鎖原子の軌道エネルギーによる方法でピーク（5 番目まで考慮する）となるアミノ酸残基から前後 2 残基ずつを候補とした。そして、お互いに共通しているものを挙げると、4 0、5 7、5 8、7 7、9 2 と 5 つがあがる。

図 2 6 により共通部分として抽出されたアミノ酸残基はいずれも実験で求められた活性に必要なアミノ酸残基（4 0、5 8、7 7、9 2）によく対応している（5 7 は 5 8 に近いために活性部位と誤って判断された。）。

[実施例 2 ; リボヌクレアーゼ A]

次に、実施例 2 の詳細について図 2 7 から図 3 0 を参照して説明する。加水分解酵素であるリボヌクレアーゼ A は実験的によく調べられていて、必須なアミノ酸残基は H i s 1 2、H i s 1 1 9 であるということが実験的に証明されている。

リボヌクレアーゼ A について、X 線結晶構造データに基づき市販のプログラム I n s i g h t I I によって水素分子を付加して、分子軌道計算に必要な座標を完成した。そして、市販のプログラム M O P A C 2 0 0 0 を用いて最適化構造を求めた後に、電子状態を得た。蛋白質まわりに水分子を置き、さらに連続誘電体近似（C O S M O 法）を用いて溶媒の影響を考慮した。

ここで、本実施例においてリボヌクレアーゼAのフロンティア軌道が分布したアミノ酸残基を図27に示す表に示す。

次に主鎖の窒素原子に着目した。図28は、本実施例2において主鎖の窒素原子に分布している分子軌道の軌道エネルギーをアミノ酸の残基番号に対して  
5 プロットした図である。本図に示すようにアミノ酸残基番号12、47、117、76、53の付近に軌道エネルギーの高い部分が現れる。また、軌道エネルギーが高いアミノ酸残基と軌道エネルギーとを抽出した表を図29に示す。軌道エネルギーの高いアミノ酸残基の周辺にあるアミノ酸残基が活性部位の候補にある。

10 また、図27に示したフロンティア軌道による候補アミノ酸残基と、図28および図29に示した主鎖原子の軌道エネルギーによる候補アミノ酸残基の共通部分を抽出した表を図30に示す。例えば、フロンティア軌道による方法で求核基の候補4つ、求電子基の候補4つを挙げた。また、主鎖原子の軌道エネルギーによる方法でピーク（5番目まで考慮する）となるアミノ酸残基から前後2残基ずつを候補とした。そして、お互いに共通しているものを挙げると、  
15 12、14、119と3つがあがる。

図30により共通部分として抽出されたアミノ酸残基はいずれも実験で求められた活性に必要なアミノ酸残基（12、119）によく対応している（14は12に近いために活性部位と誤って判断された。）。

#### 20 [他の実施の形態]

さて、これまで本発明の実施の形態について説明したが、本発明は、上述した実施の形態以外にも、上記特許請求の範囲に記載した技術的思想の範囲内において種々の異なる実施の形態にて実施されてよいものである。

例えば、蛋白質活性部位予測装置1100がスタンドアローンの形態で処理  
25 を行う場合を一例に説明したが、蛋白質活性部位予測装置1100とは別筐体で構成されるクライアント端末からの要求に応じて処理を行い、その処理結果を当該クライアント端末に返却するように構成してもよい。

また、実施形態において説明した各処理のうち、自動的に行なわれるものとして説明した処理の全部または一部を手動的に行うこともでき、あるいは、手動的に行なわれるものとして説明した処理の全部または一部を公知の方法で自動的に行うこともできる。

- 5       この他、上記文書中や図面中で示した処理手順、制御手順、具体的名称、各種の登録データや検索条件等のパラメータを含む情報、画面例、データベース構成については、特記する場合を除いて任意に変更することができる。

また、蛋白質活性部位予測装置 1 1 0 0 に関して、図示の各構成要素は機能概念的なものであり、必ずしも物理的に図示の如く構成されていることを要し  
10       ない。

例えば、蛋白質活性部位予測装置 1 1 0 0 の各部または各装置が備える処理機能、特に制御部 1 1 0 2 にて行なわれる各処理機能については、その全部または任意の一部を、CPU (Central Processing Unit) および当該CPUにて解釈実行されるプログラムにて実現することができ、  
15       あるいは、ワイヤードロジックによるハードウェアとして実現することも可能である。なお、プログラムは、後述する記録媒体に記録されており、必要に応じて蛋白質活性部位予測装置 1 1 0 0 に機械的に読み取られる。

すなわち、ROMまたはHDなどの記憶部 1 1 0 6 などには、OS (Operating System) と協働してCPUに命令を与え、各種処理を行うためのコンピュータプログラムが記録されている。このコンピュータプログラムは、RAM等にロードされることによって実行され、CPUと協働して制御部 1 1 0 2 を構成する。また、このコンピュータプログラムは、蛋白質活性部位予測装置 1 1 0 0 に対して任意のネットワーク 1 3 0 0 を介して接続されたアプリケーションプログラムサーバに記録されてもよく、必要に応じてその  
20       全部または一部をダウンロードすることも可能である。

また、本発明にかかるプログラムを、コンピュータ読み取り可能な記録媒体に格納することもできる。ここで、この「記録媒体」とは、フレキシブルディ

スク、光磁気ディスク、ROM、EPROM、EEPROM、CD-ROM、MO、DVD等の任意の「可搬用の物理媒体」や、各種コンピュータシステムに内蔵されるROM、RAM、HD等の任意の「固定用の物理媒体」、あるいは、LAN、WAN、インターネットに代表されるネットワークを介してプログラムを送信する場合の通信回線や搬送波のように、短期にプログラムを保持する「通信媒体」を含むものとする。

また、「プログラム」とは、任意の言語や記述方法にて記述されたデータ処理方法であり、ソースコードやバイナリコード等の形式を問わない。なお、「プログラム」は必ずしも単一的に構成されるものに限られず、複数のモジュールやライブラリとして分散構成されるものや、OS (Operating System) に代表される別個のプログラムと協働してその機能を達成するものをも含む。なお、実施の形態に示した各装置において記録媒体を読み取るための具体的な構成、読み取り手順、あるいは、読み取り後のインストール手順等については、周知の構成や手順を用いることができる。

記憶部 1106 に格納される各種のデータベース等（蛋白質構造データベース 1106a および処理結果データ 1106b）は、RAM、ROM等のメモリ装置、ハードディスク等の固定ディスク装置、フレキシブルディスク、光ディスク等のストレージ手段であり、各種処理やウェブサイト提供に用いる各種のプログラムやテーブルやファイルやデータベースやウェブページ用ファイル等を格納する。

また、蛋白質活性部位予測装置 1100 は、既知のパーソナルコンピュータ、ワークステーション等の情報処理端末等の情報処理装置にプリンタやモニタやイメージスキャナ等の周辺装置を接続し、該情報処理装置に本発明の方法を実現させるソフトウェア（プログラム、データ等を含む）を実装することにより実現してもよい。

さらに、蛋白質活性部位予測装置 1100 の分散・統合の具体的な形態は図示のものに限られず、その全部または一部を、各種の負荷等に応じた任意の単位

で、機能的または物理的に分散・統合して構成することができる。例えば、各データベースを独立したデータベース装置として独立に構成してもよく、また、処理の一部をCGI (Common Gateway Interface) を用いて実現してもよい。

5       また、ネットワーク1300は、蛋白質活性部位予測装置1100と外部システム1200とを相互に接続する機能を有し、例えば、インターネットや、イントラネットや、LAN (有線/無線の双方を含む) や、VANや、パソコン通信網や、公衆電話網 (アナログ/デジタルの双方を含む) や、専用回線網 (アナログ/デジタルの双方を含む) や、CATV網や、IMT2000方式、  
10   GSM方式またはPDC/PDC-P方式等の携帯回線交換網/携帯パケット交換網や、無線呼出網や、Bluetooth等の局所無線網や、PHS網や、CS、BSまたはISDB等の衛星通信網等のうちいずれかを含んでもよい。  
すなわち、本システムは、有線・無線を問わず任意のネットワークを介して、各種データを送受信することができる。

15       以上詳細に説明したように、本発明によれば、蛋白質または生理活性ポリペプチドの電子状態を分子軌道計算によって求めて、フロンティア軌道とその周辺軌道、および/または、主鎖の重原子に局在する軌道エネルギーを特定し、当該フロンティア軌道とその周辺軌道の位置、および/または、当該軌道エネルギーに基づいて当該蛋白質または生理活性ポリペプチドの活性部位となるア  
20   ミノ酸残基を予測するので、精度が高いと言われている分子軌道計算を利用し、フロンティア軌道の位置または軌道エネルギーの高い位置と反応部位との関係性を蛋白質または生理活性ポリペプチドの系に応用するため高い精度の活性部位予想を行うことができる活性部位予測装置、活性部位予測方法、プログラム、および、記録媒体を提供することができる。

25       また、本発明によれば、目的の蛋白質または生理活性ポリペプチドの構造データを取得し、取得した構造データに基づいて蛋白質または生理活性ポリペプチドの電子状態を分子軌道計算によって求めてフロンティア軌道を特定し、フ

ロンティア軌道と予め定めたエネルギー差になる分子軌道を当該フロンティア軌道の周辺軌道として決定し、フロンティア軌道と周辺軌道が分布しているアミノ酸残基を活性部位の候補のアミノ酸残基として決定し、決定された候補アミノ酸残基の中から活性部位を選択して活性部位を予測するので、精度が高い  
5 と言われている分子軌道計算を利用し、フロンティア軌道の位置と反応部位との関係性を蛋白質または生理活性ポリペプチドの系に応用するため高い精度の活性部位予想を行うことができる活性部位予測装置、活性部位予測方法、プログラム、および、記録媒体を提供することができる。

また、本発明によれば、目的の蛋白質または生理活性ポリペプチドの構造データを取得し、取得した構造データに基づいて蛋白質または生理活性ポリペプチドの電子状態を分子軌道計算によって求めて主鎖の重原子に局在する軌道エネルギーを特定し、特定された軌道エネルギーのうち予め定めた値を超える軌道エネルギーとなる分子軌道および／または相対的に軌道エネルギーが高い分子軌道が分布しているアミノ酸残基を活性部位の候補のアミノ酸残基として決定するので、精度が高いと言われている分子軌道計算を利用し、軌道エネルギーの高い位置と反応部位との関係性を蛋白質または生理活性ポリペプチドの系に応用するため高い精度の活性部位予想を行うことができる活性部位予測装置、活性部位予測方法、プログラム、および、記録媒体を提供することができる。  
10 15

また、本発明によれば、目的の蛋白質または生理活性ポリペプチドの構造データを取得し、取得した構造データに基づいて蛋白質または生理活性ポリペプチドの電子状態を分子軌道計算によって求めてフロンティア軌道を特定し、取得した構造データに基づいて蛋白質または生理活性ポリペプチドの電子状態を分子軌道計算によって求めて主鎖の重原子に局在する軌道エネルギーを特定し、フロンティア軌道と予め定めたエネルギー差になる分子軌道を当該フロンティア軌道の周辺軌道として決定し、フロンティア軌道と周辺軌道が分布しているアミノ酸残基、および／または、特定された軌道エネルギーのうち予め定めた値を超える軌道エネルギーとなる分子軌道および／または相対的に軌道エネルギー

20 25

ギーが高い分子軌道が分布しているアミノ酸残基を活性部位の候補のアミノ酸残基として決定し、決定された候補アミノ酸残基の中から活性部位を選択して活性部位を予測するので、精度が高いと言われている分子軌道計算を利用し、  
5 フロンティア軌道の位置または軌道エネルギーの高い位置と反応部位との関係性を蛋白質または生理活性ポリペプチドの系に応用するため高い精度の活性部位予想を行うことができる活性部位予測装置、活性部位予測方法、プログラム、および、記録媒体を提供することができる。

さらに、本発明によれば、分子軌道計算において、以下の3つの計算条件

- 1) 蛋白質または生理活性ポリペプチドのまわりに水分子を発生させる、
- 10 2) 蛋白質または生理活性ポリペプチドの周りに連続的な誘電体を置く、
- 3) 蛋白質または生理活性ポリペプチド表面の解離性アミノ酸残基を無電荷の状態にし、内部に埋まっている解離性アミノ酸を電荷状態にする、

のうち少なくとも一つの計算条件を設定するので、3つの計算条件を適宜設定することにより、効率的に分子軌道計算を実行することができ、さらに、活  
15 性部位予測の精度を大幅に向上させることができる活性部位予測装置、活性部位予測方法、プログラム、および、記録媒体を提供することができる。

(Ⅲ) また、以下に、本発明にかかるタンパク質相互作用情報処理装置、タンパク質相互作用情報処理方法、プログラム、および、記録媒体の実施の形態を図面に基づいて詳細に説明する。なお、この実施の形態によりこの発明が限定  
20 されるものではない。

#### [本発明の概要]

以下、本発明の概要について説明し、その後、本発明の構成および処理等について詳細に説明する。図31は本発明の基本原理を示す原理構成図である。

本発明は、概略的に、以下の基本的特徴を有する。

25 本発明は、溶媒接触面の疎水性相互作用に基づき不安定性の高い部位を特定する。すなわち、本発明は、まず、相互作用する複数のタンパク質の単体時の溶媒接触面積（溶媒分子が接することのできる分子表面の面積であり「溶媒露

出表面積」ともいう)と複合体形成時の溶媒接触面積とを別々に計算し、その差を取ることで相互作用部分の溶媒接触面を求める。つまり、単体時の溶媒接触面積と複合体形成時の溶媒接触面積との差が大きい部位は、複合体を形成すると溶媒に接触する面積が少なくなること示すため、このような部位は相互作用部位である可能性が高いといえるので、差が大きいアミノ酸残基部位を相互作用部位の溶媒接触面として特定する。なお、複合体時の構造データがない場合本処理は行わない。

そして、本発明はタンパク質の溶媒接触面について疎水性相互作用エネルギーを求めることにより、タンパク質の一次構造を構成するアミノ酸残基中で溶媒面でありかつ疎水面である部位を特定する。このような部位は、単体時には不安定性が高く、複合体形成時には疎水面が覆い隠されることにより安定化すると考えられるため、相互作用部位となる可能性が高い部位であるといえる。

また、本発明は、タンパク質中において静電相互作用エネルギーの高い部位を特定することにより不安定性の高い部位を特定する。すなわち、本発明は、分子軌道法などにより求められた原子電荷(部分電荷)に基づいて静電相互作用エネルギーの高い部位を計算する。このような部位は単体時には不安定性が高く、複合体形成時にはエネルギー的に安定化すると考えられるため、相互作用部位となる可能性が高い部位であるといえる。ここで、原子電荷は分子軌道法などによる各種の計算手法により求めてもよいし、分子力学法や分子動力学法等において各種のパラメータ値として与えられた原子電荷の値などを流用してもよい。

そして、本発明は、溶媒接触面、疎水性相互作用エネルギー、および、静電相互作用エネルギーに基づいて不安定性の高い部位を特定することにより相互作用部位を特定する。

[システム構成]



まず、本システムの構成について説明する。図 3 2 は、本発明が適用される本システムの構成の一例を示すブロック図であり、該構成のうち本発明に係る部分のみを概念的に示している。本システムは、概略的に、タンパク質相互作用情報処理装置 2 1 0 0 と、配列情報等に関する外部データベースやホモロジー検索等の外部プログラム等を提供する外部システム 2 2 0 0 とを、ネットワーク 2 3 0 0 を介して通信可能に接続して構成されている。

図 3 2 においてネットワーク 2 3 0 0 は、タンパク質相互作用情報処理装置 2 1 0 0 と外部システム 2 2 0 0 とを相互に接続する機能を有し、例えば、インターネット等である。

図 3 2 において外部システム 2 2 0 0 は、ネットワーク 2 3 0 0 を介して、タンパク質相互作用情報処理装置 2 1 0 0 と相互に接続され、利用者に対して DNA 等の配列情報やタンパク質等の構造情報等に関する外部データベースやホモロジー検索やモチーフ検索等の外部プログラムを実行するウェブサイトを提供する機能を有する。

ここで、外部システム 2 2 0 0 は、WEB サーバや ASP サーバ等として構成してもよく、そのハードウェア構成は、一般に市販されるワークステーション、パーソナルコンピュータ等の情報処理装置およびその付属装置により構成してもよい。また、外部システム 2 2 0 0 の各機能は、外部システム 2 2 0 0 のハードウェア構成中の CPU、ディスク装置、メモリ装置、入力装置、出力装置、通信制御装置等およびそれらを制御するプログラム等により実現される。

図 3 2 においてタンパク質相互作用情報処理装置 2 1 0 0 は、概略的に、タンパク質相互作用情報処理装置 2 1 0 0 の全体を統括的に制御する CPU 等の制御部 2 1 0 2、通信回線等に接続されるルータ等の通信装置（図示せず）に接続される通信制御インターフェース部 2 1 0 4、入力装置 2 1 1 2 や出力装置 2 1 1 4 に接続される入出力制御インターフェース部 2 1 0 8、および、各種のデータベースやテーブルなどを格納する記憶部 2 1 0 6 を備えて構成されており、これら各部は任意の通信路を介して通信可能に接続されている。さら

に、このタンパク質相互作用情報処理装置 2 1 0 0 は、ルータ等の通信装置および専用線等の有線または無線の通信回線を介して、ネットワーク 2 3 0 0 に通信可能に接続されている。

記憶部 2 1 0 6 に格納される各種のデータベースやテーブル（タンパク質構造データベース 2 1 0 6 a および処理結果データ 2 1 0 6 b）は、固定ディスク装置等のストレージ手段であり、各種処理に用いる各種のプログラムやテーブルやファイルやデータベースやウェブページ用ファイル等を格納する。

これら記憶部 2 1 0 6 の各構成要素のうち、タンパク質構造データベース 2 1 0 6 a は、タンパク質のアミノ酸配列情報（一次構造データ）や立体構造データ（構成原子の 3 次元座標データ等）や各種のアノテーション情報を格納したデータベースである。タンパク質構造データベース 2 1 0 6 a は、インターネットを経由してアクセスする外部のデータベースであってもよく、また、これらのデータベースをコピーしたり、オリジナルの配列情報を格納したり、さらに独自のアノテーション情報等を付加したりして作成したインハウスデータベースであってもよい。

また、処理結果データ 2 1 0 6 b は、処理結果データに関する情報等を格納する処理結果データ格納手段である。

また、図 3 2 において、通信制御インターフェース部 2 1 0 4 は、タンパク質相互作用情報処理装置 2 1 0 0 とネットワーク 2 3 0 0（またはルータ等の通信装置）との間における通信制御を行う。すなわち、通信制御インターフェース部 2 1 0 4 は、他の端末と通信回線を介してデータを通信する機能を有する。

また、図 3 2 において、入出力制御インターフェース部 2 1 0 8 は、入力装置 2 1 1 2 や出力装置 2 1 1 4 の制御を行う。ここで、出力装置 2 1 1 4 としては、モニタ（家庭用テレビを含む）の他、スピーカを用いることができる（なお、以下においては出力装置 2 1 1 4 をモニタとして記載する場合がある）。また、入力装置 2 1 1 2 としては、キーボード、マウス、および、マイ

ク等を用いることができる。また、モニタも、マウスと協働してポインティングデバイス機能を実現する。

また、図 3 2 において、制御部 2 1 0 2 は、OS (O p e r a t i n g S y s t e m) 等の制御プログラム、各種の処理手順等を規定したプログラム、  
5 および所要データを格納するための内部メモリを有し、これらのプログラム等により、種々の処理を実行するための情報処理を行う。制御部 2 1 0 2 は、機能概念的に、構造データ取得部 2 1 0 2 a、溶媒接触面特定部 2 1 0 2 b、疎水面特定部 2 1 0 2 c、静電相互作用部位特定部 2 1 0 2 d、相互作用部位特定部 2 1 0 2 e、および、相互作用予測部 2 1 0 2 f を備えて構成されている。

10 このうち、構造データ取得部 2 1 0 2 a は、相互作用する複数のタンパク質の一次構造データと単体時および／または複合体形成時の立体構造データとを含む構造データを取得する構造データ取得手段である。また、溶媒接触面特定部 2 1 0 2 b は、構造データ取得手段にて取得した構造データに従って、一次構造データを構成するアミノ酸残基毎に溶媒接触面を特定する溶媒接触面特定  
15 手段である。

また、疎水面特定部 2 1 0 2 c は、構造データ取得手段にて取得した構造データに従って、一次構造データを構成するアミノ酸残基毎に疎水性相互作用エネルギーを特定する疎水面特定手段である。また、静電相互作用部位特定部 2 1 0 2 d は、構造データ取得手段にて取得した構造データに従って、一次構造  
20 データを構成するアミノ酸残基毎に静電相互作用エネルギーを特定する静電相互作用部位特定手段である。

また、相互作用部位特定部 2 1 0 2 e は、溶媒接触面特定手段にて特定された溶媒接触面、疎水面特定手段にて特定された疎水性相互作用エネルギー、および、静電相互作用部位特定手段にて特定された静電相互作用エネルギーに基づいて不安定性の高いアミノ酸残基の部位を特定することにより相互作用部位  
25 を特定する相互作用部位特定手段である。

また、相互作用予測部 2 1 0 2 f は、相互作用部位特定手段にて特定された相互作用部位について、相互作用する相手側の一次配列を特定し、当該一次配列を含む一次構造を持つ候補タンパク質を検索する候補タンパク質検索部 2 1 0 2 g を備え、候補タンパク質検索手段にて検索された候補タンパク質について、構造データ取得手段、溶媒接触面特定手段、疎水面特定手段、静電相互作用部位特定手段、および、相互作用部位特定手段を実行することにより、相手側の一次配列の部分が候補タンパク質の相互作用部位として特定されるか否かを確認する手段である。なお、これら各部によって行なわれる処理の詳細については、後述する。

#### 10 [システムの処理]

次に、このように構成された本実施の形態における本システムの処理の一例について、以下に図 3 3 ～図 4 2 を参照して詳細に説明する。

#### [メイン処理]

まず、メイン処理の詳細について図 3 3 を参照して説明する。図 3 3 は、本  
15 実施形態における本システムのメイン処理の一例を示すフローチャートである。

タンパク質相互作用情報処理装置 2 1 0 0 は、構造データ取得部 2 1 0 2 a の処理により、タンパク質構造データベース 2 1 0 6 a または外部システム 2 2 0 0 の外部データベース（例えば、PDB（P r o t e i n D a t a B a n k）など）にアクセスして、相互作用する複数のタンパク質の一次構造データと単体時および／または複合体形成時の立体構造データとを含む構造データ  
20 を取得する（ステップ S A 2 - 1）。ここで、取得する構造データは、相互作用する複数のタンパク質の単独時の構造データと複合体形成時の構造データの双方を含む場合と、相互作用する複数のタンパク質の単独時の構造データのみの場合を含む。

25 ついで、タンパク質相互作用情報処理装置 2 1 0 0 は、複合体時の構造データがある場合には溶媒接触面特定部 2 1 0 2 b の処理により、図 3 4 を用いて後述するように、相互作用する複数のタンパク質の単独時の構造データと複合

体形成時の構造データの双方に従って、一次構造データを構成するアミノ酸残基毎に溶媒接触面を特定する（ステップSA2-2）。ここで、溶媒接触面特定処理の詳細について図34を参照して説明する。図34は、本実施形態における本システムの溶媒接触面特定処理の一例を示すフローチャートである。

- 5       まず、溶媒接触面特定部2102bは、各残基について、単体時の溶媒接触面積 $S_{\text{isolated}}$ を計算する（ステップSB2-1）。ここで、本発明において溶媒接触面積を求める手法については既知のいずれの手法を用いてもよく、例えば、文献1（“Numerical Calculation of Molecular Surface Area. I. Assessment of Errors” A. A. Bliznyuk and J. E. Gready, J. Comput. Chem., 17, 962-969 (1996).）や、文献2（“Numerical Calculation of Molecular Surface Area. II. Assessment of Errors” A. A. Bliznyuk and J. E. Gready, J. Comput. Chem., 17, 970-975 (1996).）に開示された手法を用いてもよい。

ついで、溶媒接触面特定部2102bは、各残基について、複合体形成時の溶媒接触面積 $S_{\text{complex}}$ を計算する（ステップSB2-2）。

- 20       ついで、溶媒接触面特定部2102bは、各残基について、単体時の溶媒接触面積 $S_{\text{isolated}}$ と、複合体形成時の溶媒接触面積 $S_{\text{complex}}$ の差分を計算する（ステップSB2-3）。これにて、溶媒接触面特定処理が終了する。

- 再び図33に戻り、タンパク質相互作用情報処理装置2100は、疎水面特定部2102cの処理により、図35を用いて後述するように、相互作用する複数のタンパク質の単独時の構造データと複合体形成時の構造データの双方に従って、タンパク質の一次構造を構成する各アミノ酸残基および原子毎の疎水性パラメータなどに基づいて、各残基および原子毎に疎水性相互作用エネルギーを計算し疎水面を特定する（ステップSA2-3）。例えば、アミノ酸残基

が  $Lys$  の場合は、 $\epsilon$  位の窒素  $N$  とそれに結合している水素  $H$  は親水的とし、 $\beta$ 、 $\gamma$ 、 $\delta$  位の炭素  $C$  とそれに結合している水素  $H$  は疎水的とする。

ここで、疎水面特定処理の詳細について図 35 を参照して説明する。図 35 は、本実施形態における本システムの疎水面特定処理の一例を示すフローチャートである。本例では、タンパク質  $A$  とタンパク質  $B$  とが相互作用する場合を一例に説明する。

まず、疎水面特定部 2102c は、数式 1 により疎水面の減少量を計算する (ステップ SC2-1)。

$$\Delta S_{\text{hydrophobic}} = S_{\text{hydrophobicA}} + S_{\text{hydrophobicB}} - S_{\text{hydrophobicAB}} \cdots \text{数式 1}$$

ここで、 $\Delta S_{\text{hydrophobic}}$  は疎水面の減少量であり、 $S_{\text{hydrophobicA}}$  はタンパク質  $A$  単体時の疎水面の面積であり、 $S_{\text{hydrophobicB}}$  はタンパク質  $B$  単体時の疎水面の面積であり、 $S_{\text{hydrophobicAB}}$  はタンパク質  $A$  とタンパク質  $B$  が複合体を形成した時の疎水面の面積である。

そして、疎水面特定部 2102c は、数式 2 により疎水性相互作用エネルギー  $E_{\text{hydrophobic}}$  を計算する (SC2-2)。

$$E_{\text{hydrophobic}} = k \times \Delta S_{\text{hydrophobic}} \cdots \text{数式 2}$$

20

ここで、 $k = 24 \text{ cal/mol} \cdot \text{\AA}^2$  である

(参考 "Quantification of the hydrophobic interaction by simulations of the aggregation of small hydrophobic solutions in water", T. M. Raschke, J. Tsai and M. Levitt, PNAS, 98, 5965-5

9 6 9 (2 0 0 1) )。

そして、疎水面特定部 2 1 0 2 c は、予め定めた閾値を超える疎水性相互作用エネルギーであるアミノ酸残基部位を疎水面として特定する（ステップ S C 2-3）。これにて、疎水面特定処理が終了する。

5 再び図 3 3 に戻り、タンパク質相互作用情報処理装置 2 1 0 0 は、静電相互作用部位特定部 2 1 0 2 d の処理により、図 3 6 を用いて後述するように、相互作用する複数のタンパク質の単独時の構造データと複合体形成時の構造データの双方に従って、上記一次構造データを構成するアミノ酸残基毎に静電相互作用エネルギーを特定する（ステップ S A 2-4）。ここで、静電相互作用部位特定処理の詳細について図 3 6 を参照して説明する。図 3 6 は、本実施形態における本システムの静電相互作用部位特定処理の一例を示すフローチャートである。

まず、静電相互作用部位特定部 2 1 0 2 d は、数式 3 により各残基について静電相互作用エネルギー  $E_n$  を計算する（ステップ S D 2-1）。

$$15 \quad E_n = \frac{1}{4\pi\epsilon} \sum_{i \in n} \sum_{j \notin n} \frac{q_i q_j}{R_{ij}} \quad \cdot \cdot \cdot \text{数式 3}$$

ここで、 $\epsilon$  は分子内部における誘電率であり、 $q$  は部分電荷であり、 $i$  と  $j$  は原子を示す添え字であり、 $R$  は原子  $i$  と原子  $j$  の間の距離である。 $E_n$  は静電相互作用であり、分子内部の極性部位およびイオン化して荷電している部位間の相互作用を原子核上に部分電荷を置くことで近似している。これにて、静電相互作用部位特定処理が終了する。

20 再び図 3 3 に戻り、タンパク質相互作用情報処理装置 2 1 0 0 は、相互作用部位特定部 2 1 0 2 e の処理により、図 3 7 を用いて後述するように、溶媒接触面、疎水性相互作用エネルギー、および、静電相互作用エネルギーに基づいて不安定性の高い上記アミノ酸残基の部位を特定することにより相互作用部位を特定する（ステップ S A 2-5）。ここで、相互作用部位特定処理の詳細に

ついて図 3 7 を参照して説明する。図 3 7 は、本実施形態における本システムの相互作用部位特定処理の一例を示すフローチャートである。

まず、相互作用部位特定部 2 1 0 2 e は、溶媒接触面積の差  $\Delta S$  が予め定めた閾値を超える部位を特定する（ステップ S E 2 - 1）。

- 5      次に、相互作用部位特定部 2 1 0 2 e は、疎水性相互作用エネルギー  $E_{\text{hydrophobic}}$  が予め定めた閾値を超える部位を特定する（ステップ S E 2 - 2）。

次に、相互作用部位特定部 2 1 0 2 e は、静電相互作用エネルギー  $E_n$  が予め定めた閾値を超える部位を特定する（ステップ S E 2 - 3）。これにて、相互作用部位特定処理が終了する。これにて、メイン処理が終了する。

- 10      [相互作用部位予測処理]

次に、相互作用部位予測処理の詳細について図 3 8 を参照して説明する。図 3 8 は、本実施形態における本システムの相互作用部位予測処理の一例を示すフローチャートである。

- 15      まず、タンパク質相互作用情報処理装置 2 1 0 0 は、上述したメイン処理により相互作用部位を特定する（ステップ S F 2 - 1）。

- 20      ついで、相互作用予測部 2 1 0 2 f は、候補タンパク質検索部 2 1 0 2 g の処理により、ステップ S F 2 - 1 で特定した相互作用部位と相互作用する相手側の一次配列（同一タンパク質内の配列を含む）を特定し（ステップ S F 2 - 2）、当該一次配列を含む一次構造を持つ候補タンパク質を検索する（ステップ S F 2 - 3）。

- 25      ついで、相互作用予測部 2 1 0 2 f は、候補タンパク質について、上述した構造データ取得処理、溶媒接触面特定処理（複合体時の構造データがある場合）、疎水面特定処理、静電相互作用部位特定処理、および、相互作用部位特定処理を実行することにより、相手側の一次配列の部分が候補タンパク質の相互作用部位として特定されるか否かを確認する（ステップ S F 2 - 4）。これにて、相互作用部位予測処理が終了する。

[実施例 1]



次に、実施例 1 の詳細について図 3 9 ～図 4 4 を参照して説明する。実施例 1 は、タンパク質として「b a r n a s e」と「b a r s t a r」とを用いて相互作用部位を特定する場合を一例に説明する。

図 3 9 は、タンパク質相互作用情報処理装置 1 0 0 が、溶媒接触面特定部 1 0 2 b の処理により、b a r n a s e－b a r s t a r 複合体の結晶構造に基づき、b a r n a s e について各アミノ酸残基毎に溶媒接触面積の差  $\Delta S$  を計算した処理図である。本図に示すように、b a r n a s e の一次構造のうち、3 8 番目、5 9 番目、8 3 番目、および、1 0 2 番目のアミノ酸残基の差  $\Delta S$  が大きく、b a r n a s e はこの部位で b a r s t a r と相互作用していることが特定できる。

また、図 4 0 は、タンパク質相互作用情報処理装置 1 0 0 が、疎水面特定部 1 0 2 c の処理により、b a r n a s e 単体の結晶構造に基づき、b a r n a s e について各アミノ酸残基毎に疎水性相互作用エネルギーを計算した処理図である。本図に示すように、8 2 番目のアミノ酸残基の疎水相互作用エネルギーが高く、この付近で相互作用の可能性があることを示している。

また、図 4 1 は、タンパク質相互作用情報処理装置 1 0 0 が、静電相互作用特定部 1 0 2 d の処理により、b a r n a s e 単体の結晶構造に基づき、b a r n a s e について各アミノ酸残基毎に静電相互作用エネルギーを計算した処理図である。本図に示すように、5 9 番目、6 6 番目、8 3 番目、および、1 0 2 番目のアミノ酸残基の静電相互作用エネルギーが高く、この付近で相互作用の可能性があることを示している。

また、図 4 2 は、タンパク質相互作用情報処理装置 1 0 0 が、溶媒接触面特定部 1 0 2 b の処理により、b a r n a s e－b a r s t a r 複合体の結晶構造に基づき、b a r s t a r について各アミノ酸残基毎に溶媒接触面積の差  $\Delta S$  を計算した処理図である。本図に示すように、b a r s t a r の一次構造のうち、3 0 番目、3 6 番目、4 0 番目、4 5 番目、4 7 番目、および、7 7 番

目のアミノ酸残基の差 $\Delta S$ が大きく、`bars tar`はこの部位で`bar n a s e`と相互作用していることが特定できる。

また、図43は、タンパク質相互作用情報処理装置100が、疎水面特定部102cの処理により、`bars tar`単体の結晶構造に基づき、`bars t a r`について各アミノ酸残基毎に疎水性相互作用エネルギーを計算した処理図  
5 である。本図に示すように、30番目のアミノ酸残基の疎水相互作用エネルギーが高く、この付近で相互作用の可能性があることを示している。

また、図44は、タンパク質相互作用情報処理装置100が、静電相互作用特定部102dの処理により、`bars tar`単体の結晶構造に基づき、`b a r s t a r`について各アミノ酸残基毎に静電相互作用エネルギーを計算した処理  
10 図である。本図に示すように、35番目、39番目、58番目、65番目、77番目、および、80番目のアミノ酸残基の静電相互作用エネルギーが高く、この付近で相互作用の可能性があることを示している。

タンパク質相互作用情報処理装置100は、相互作用作用部位特定部102  
15 eの処理により、図40、図41に示す結果に基づいて、`bar n a s e`については59番目、66番目、82番目、83番目、および、102番目アミノ酸残基残基が相互作用候補部位であると特定する。このことは、図39に示す相互作用部位の複合体における既知情報の結果と良く一致しており、タンパク質単体構造から複合体形成時における結合部位を予測できることを示している。

また、タンパク質相互作用情報処理装置100は、相互作用作用部位特定部1  
20 02eの処理により、図43、図44に示す結果に基づいて、`bars tar`については30番目、35番目、39番目、58番目、65番目、77番目、および、80番目のアミノ酸残基残基が相互作用候補部位であると特定する。このことは、図42に示す相互作用部位の複合体における既知情報の結果と良  
25 く一致しており、同様に、タンパク質単体構造から複合体形成時における結合部位を予測できることを示している。これにて、実施例1が終了する。

[実施例2]

次に、実施例 2 の詳細について図 4 5 ～図 5 0 を参照して説明する。実施例 2 は、タンパク質として *Ribonuclease* とその *Inhibitor* とを用いて相互作用部位を特定する場合を一例に説明する。

図 4 5 は、タンパク質相互作用情報処理装置 1 0 0 が、溶媒接触面特定部 1 0 2 b の処理により、*Ribonuclease-inhibitor* 複合体の結晶構造に基づき、*Ribonuclease* について各アミノ酸残基毎に溶媒接触面積の差  $\Delta S$  を計算した処理図である。本図に示すように、*Ribonuclease* の一次構造のうち、3 9 番目のアミノ酸残基の差  $\Delta S$  が大きく、*Ribonuclease* はこの部位で *inhibitor* と相互作用していることが特定できる。

また、図 4 6 は、タンパク質相互作用情報処理装置 1 0 0 が、疎水面特定部 1 0 2 c の処理により、*Ribonuclease* 単体の結晶構造に基づき、*Ribonuclease* について各アミノ酸残基毎に疎水性相互作用エネルギーを計算した処理図である。本図に示すように、疎水性相互作用エネルギーに関しては際立ったピークは判別できない。

また、図 4 7 は、タンパク質相互作用情報処理装置 1 0 0 が、静電相互作用特定部 1 0 2 d の処理により、*Ribonuclease* 単体の結晶構造に基づき、*Ribonuclease* について各アミノ酸残基毎に静電相互作用エネルギーを計算した処理図である。本図に示すように、1 番目、7 番目、および、3 9 番目のアミノ酸残基の静電相互作用エネルギーが高く、この付近で相互作用の可能性があることを示している。

また、図 4 8 は、タンパク質相互作用情報処理装置 1 0 0 が、溶媒接触面特定部 1 0 2 b の処理により、*Ribonuclease-inhibitor* 複合体の結晶構造に基づき、*inhibitor* について各アミノ酸残基毎に溶媒接触面積の差  $\Delta S$  を計算した処理図である。本図に示すように、*inhibitor* の一次構造のうち、4 3 3 番目のアミノ酸残基の差  $\Delta S$  が大きく、

i n h i b i t o rはこの部位でR i b o n u c l e a s eと相互作用していることが特定できる。

また、図49は、タンパク質相互作用情報処理装置100が、疎水面特定部102cの処理により、i n h i b i t o r単体の結晶構造に基づき、i n h i b i t o rについて各アミノ酸残基毎に疎水性相互作用エネルギーを計算した処理図である。本図に示すように、433番目のアミノ酸残基の疎水相互作用エネルギーが高く、この付近で相互作用の可能性があることを示している。

また、図50は、タンパク質相互作用情報処理装置100が、静電相互作用特定部102dの処理により、i n h i b i t o r単体の結晶構造に基づき、i n h i b i t o rについて各アミノ酸残基毎に静電相互作用エネルギーを計算した処理図である。本図に示すように、433番目付近のアミノ酸残基の静電相互作用エネルギーが高く、この付近で相互作用の可能性があることを示している。

タンパク質相互作用情報処理装置100は、相互作用作用部位特定部102eの処理により、図46、図47に示す結果に基づいて、R i b o n u c l e a s eについては1番目、7番目、および、39番目アミノ酸残基残基が相互作用候補部位であると特定する。このことは、図45に示す相互作用部位の複合体における既知情報の結果と良く一致しており、タンパク質単体構造から複合体形成時における結合部位を予測できることを示している。また、タンパク質相互作用情報処理装置100は、相互作用作用部位特定部102eの処理により、図49、図50に示す結果に基づいて、i n h i b i t o rについては433番目のアミノ酸残基残基が相互作用候補部位であると特定する。このことは、図48に示す相互作用部位の複合体における既知情報の結果と良く一致しており、同様に、タンパク質単体構造から複合体形成時における結合部位を予測できることを示している。これにて、実施例2が終了する。

[他の実施の形態]

さて、これまで本発明の実施の形態について説明したが、本発明は、上述した実施の形態以外にも、上記特許請求の範囲に記載した技術的思想の範囲内において種々の異なる実施の形態にて実施されてよいものである。

5 本実施例においては、相互作用するタンパク質の単体時および複合体時の構造データを用いて溶媒接触面を特定した結果と、単体時の構造データを用いて疎水性相互作用および静電相互作用を求めた結果との間に相関があることを示している。しかしながら、単体時の構造データのみを用いて疎水性相互作用および静電相互作用を求めた場合であっても本発明の効果をを得ることができることは自明である。

10 また、例えば、タンパク質相互作用情報処理装置 2 1 0 0 がスタンドアローンの形態で処理を行う場合を一例に説明したが、タンパク質相互作用情報処理装置 2 1 0 0 とは別筐体で構成されるクライアント端末からの要求に応じて処理を行い、その処理結果を当該クライアント端末に返却するように構成してもよい。

15 また、実施形態において説明した各処理のうち、自動的に行なわれるものとして説明した処理の全部または一部を手動的に行うこともでき、あるいは、手動的に行なわれるものとして説明した処理の全部または一部を公知の方法で自動的に行うこともできる。

20 この他、上記文書中や図面中で示した処理手順、制御手順、具体的名称、各種の登録データや検索条件等のパラメータを含む情報、画面例、データベース構成については、特記する場合を除いて任意に変更することができる。

また、タンパク質相互作用情報処理装置 2 1 0 0 に関して、図示の各構成要素は機能概念的なものであり、必ずしも物理的に図示の如く構成されていることを要しない。

25 例えば、タンパク質相互作用情報処理装置 2 1 0 0 の各部または各装置が備える処理機能、特に制御部 2 1 0 2 にて行なわれる各処理機能については、その全部または任意の一部を、CPU (C e n t r a l P r o c e s s i n g

Unit) および当該CPUにて解釈実行されるプログラムにて実現することができ、あるいは、ワイヤードロジックによるハードウェアとして実現することも可能である。なお、プログラムは、後述する記録媒体に記録されており、必要に応じてタンパク質相互作用情報処理装置2100に機械的に読み取られる。

すなわち、ROMまたはHDなどの記憶部2106などには、OS (Operating System) と協働してCPUに命令を与え、各種処理を行うためのコンピュータプログラムが記録されている。このコンピュータプログラムは、RAM等にロードされることによって実行され、CPUと協働して制御部2102を構成する。また、このコンピュータプログラムは、タンパク質相互作用情報処理装置2100に対して任意のネットワーク2300を介して接続されたアプリケーションプログラムサーバに記録されてもよく、必要に応じてその全部または一部をダウンロードすることも可能である。

また、本発明にかかるプログラムを、コンピュータ読み取り可能な記録媒体に格納することもできる。ここで、この「記録媒体」とは、フレキシブルディスク、光磁気ディスク、ROM、EPROM、EEPROM、CD-ROM、MO、DVD等の任意の「可搬用の物理媒体」や、各種コンピュータシステムに内蔵されるROM、RAM、HD等の任意の「固定用の物理媒体」、あるいは、LAN、WAN、インターネットに代表されるネットワークを介してプログラムを送信する場合の通信回線や搬送波のように、短期にプログラムを保持する「通信媒体」を含むものとする。

また、「プログラム」とは、任意の言語や記述方法にて記述されたデータ処理方法であり、ソースコードやバイナリコード等の形式を問わない。なお、「プログラム」は必ずしも単一的に構成されるものに限られず、複数のモジュールやライブラリとして分散構成されるものや、OS (Operating System) に代表される別個のプログラムと協働してその機能を達成するものをも含む。なお、実施の形態に示した各装置において記録媒体を読み取る

ための具体的な構成、読み取り手順、あるいは、読み取り後のインストール手順等については、周知の構成や手順を用いることができる。

記憶部 2 1 0 6 に格納される各種のデータベース等（タンパク質構造データベース 2 1 0 6 a および処理結果データ 2 1 0 6 b）は、RAM、ROM等の  
5 メモリ装置、ハードディスク等の固定ディスク装置、フレキシブルディスク、光ディスク等のストレージ手段であり、各種処理やウェブサイト提供に用いる各種のプログラムやテーブルやファイルやデータベースやウェブページ用ファイル等を格納する。

また、タンパク質相互作用情報処理装置 2 1 0 0 は、既知のパーソナルコン  
10 ピュータ、ワークステーション等の情報処理端末等の情報処理装置にプリンタやモニタやイメージスキャナ等の周辺装置を接続し、該情報処理装置に本発明の方法を実現させるソフトウェア（プログラム、データ等を含む）を実装することにより実現してもよい。

さらに、タンパク質相互作用情報処理装置 2 1 0 0 の分散・統合の具体的形態は図示のものに限られず、その全部または一部を、各種の負荷等に応じた任意の単位で、機能的または物理的に分散・統合して構成することができる。例えば、各データベースを独立したデータベース装置として独立に構成してもよく、また、処理の一部をCGI（Common Gateway Interface）を用いて実現してもよい。

また、ネットワーク 2 3 0 0 は、タンパク質相互作用情報処理装置 2 1 0 0 と外部システム 2 2 0 0 とを相互に接続する機能を有し、例えば、インターネットや、イントラネットや、LAN（有線／無線の双方を含む）や、VANや、パソコン通信網や、公衆電話網（アナログ／デジタルの双方を含む）や、専用回線網（アナログ／デジタルの双方を含む）や、CATV網や、IMT 2 0 0  
20 0 方式、GSM方式またはPDC／PDC-P方式等の携帯回線交換網／携帯パケット交換網や、無線呼出網や、Bluetooth等の局所無線網や、PHS網や、CS、BSまたはISDB等の衛星通信網等のうちいずれかを含ん

でもよい。すなわち、本システムは、有線・無線を問わず任意のネットワークを介して、各種データを送受信することができる。

以上詳細に説明したように、本発明によれば、相互作用する複数のタンパク質の一次構造データと単体時および／または複合体形成時の立体構造データとを含む構造データを取得し、取得した構造データに従って、一次構造データを構成するアミノ酸残基毎に疎水性相互作用エネルギーを特定し、取得した構造データに従って、一次構造データを構成するアミノ酸残基毎に静電相互作用エネルギーを特定し、特定された疎水性相互作用エネルギー、および、静電相互作用エネルギーに基づいて不安定性の高いアミノ酸残基の部位を特定することにより相互作用部位を特定するので、構造データによりタンパク質の相互作用部位を容易に特定することができるタンパク質相互作用情報処理装置、タンパク質相互作用情報処理方法、プログラム、および、記録媒体を提供することができる。

また、本発明によれば、取得した構造データに従って、一次構造データを構成するアミノ酸残基毎に溶媒接触面を特定し、特定された溶媒接触面、疎水性相互作用エネルギー、および、静電相互作用エネルギーに基づいて不安定性の高いアミノ酸残基の部位を特定することにより相互作用部位を特定するので、複合体時の構造データがある場合に、タンパク質の相互作用部位をさらに正確かつ容易に特定することができるタンパク質相互作用情報処理装置、タンパク質相互作用情報処理方法、プログラム、および、記録媒体を提供することができる。

さらに、本発明によれば、相互作用部位特定手段にて特定された相互作用部位について、相互作用する相手側の一次配列を特定し、当該一次配列を含む一次構造を持つ候補タンパク質を検索し、検索された候補タンパク質について、上述した構造データ取得手段、溶媒接触面特定手段（複合体時の構造データがある場合）、疎水面特定手段、静電相互作用部位特定手段、および、相互作用部位特定手段を実行することにより、相手側の一次配列の部分が候補タンパク



質の相互作用部位として特定されるか否かを確認するので、未知の相互作用を予測することが容易にできるタンパク質相互作用情報処理装置、タンパク質相互作用情報処理方法、プログラム、および、記録媒体を提供することができる。

(IV) また、以下に、本発明にかかる結合部位予測装置、結合部位予測方法、  
5 プログラム、および、記録媒体の実施の形態を図面に基づいて詳細に説明する。  
なお、この実施の形態によりこの発明が限定されるものではない。

特に、以下の実施の形態においては、本発明を、蛋白質のアミノ酸配列等に適用した例について説明するが、この場合に限られず、生理活性ポリペプチドのアミノ酸配列を用いた場合においても同様に適用することができる。

#### 10 [本発明の概要]

以下、本発明の概要について説明し、その後、本発明の構成および処理等について詳細に説明する。図5 1および図5 2は本発明の基本原理を示す原理構成図である。本発明は、概略的に、以下の基本的特徴を有する。

図5 1は、本発明により1つの蛋白質のアミノ酸配列情報からその蛋白質の  
15 結合部位を予測する場合の概念を説明する図である。

図5 1に示すように、まず、本発明では、蛋白質または生理活性ポリペプチドのアミノ酸配列データから蛋白質の立体構造における各アミノ酸残基間の空間上の距離データを求める（ステップS A 3-1）。

ここで、空間上の距離データを求める手法には、例えば、以下の3通りの手法がある。  
20

##### 1) 高速計算手法

この手法は、アミノ酸間の配列上の距離から空間的な距離に変換する方法である。図5 6は、本発明の高速計算手法の概念を示す図である。蛋白質の立体構造がガウス鎖をとると仮定すると、蛋白質のアミノ酸配列上の距離と、蛋白質の立体構造における空間上の距離は以下の式で関係づけることができる。  
25

$$r = k d^n \quad (0 < n < 1)$$

ここで  $r$  は空間上の距離、 $d$  は配列上の距離、 $k$  は比例定数である。つまり、配列上の距離  $d$  が分かれば、空間上の距離  $r$  を計算することができる。 $k$  および  $n$  の値は、例えば、PDB (Protein Data Bank) などの  
 5 蛋白質構造データベースに収集された立体構造情報データからアミノ酸間の配列上の距離と空間上の距離との関係を統計学的に処理し適切な値を設定してもよい。ここで、 $n$  は、0～1であるが、好ましくは、0.5～0.6である。また、 $k$  は2.8 Å～4.8 Åであるが、好ましくは、3.3 Å～4.3 Åである。本手法はアルゴリズムが非常にシンプルであり計算負荷も非常に少ない  
 10 ため、例えば数万以上の蛋白質を扱う時などの大量の蛋白質について処理する場合には非常に強力な方法となる。

## 2) 構造データ利用計算手法

この手法は、蛋白質構造データベースに登録された立体構造情報データを用いることにより、正確に実際のアミノ酸残基間の空間上の距離を求める手法である。例えばPDBなどの蛋白質構造データベースに目的の蛋白質の立体構造  
 15 情報データが格納されている場合には、データベースに登録された立体構造情報データを取得することにより、空間上の距離を以下のように正確に計算することができる。

例えばアミノ酸残基番号  $I$  とアミノ酸残基番号  $J$  との空間上の距離  $R_{IJ}$  は、  
 20 アミノ酸残基番号  $I$  の重心や特定の主鎖の原子などの座標を ( $x_I$ ,  $y_I$ ,  $z_I$ ) とし、アミノ酸残基番号  $J$  の重心や特定の主鎖の原子などの座標を ( $x_J$ ,  $y_J$ ,  $z_J$ ) とすると、以下のように計算することができる。

$$R_{IJ}^2 = (x_I - x_J)^2 + (y_I - y_J)^2 + (z_I - z_J)^2$$

(ここで、 $R_{IJ} > 0$ )

### 3) シミュレーションデータ利用計算手法

この手法は、構造未知の蛋白質について、既知の構造シミュレーション手法により蛋白質の構造シミュレーションを行い、そのシミュレーションデータ

- 5 (予想される立体構造情報データ) を用いて空間上の距離を求める手法である。立体構造予測シミュレーション手法に関してはホモロジー・モデリングなど様々な方法がある。例えば、『実践バイオインフォマティクス』(C. G i b a s, P. J a m b e c k 著 オライリー・ジャパン 2002年)などに詳しく紹介されている。

- 10 本手法は、手法1や手法2と比較して計算負荷が大きいという欠点があるが、構造未知の蛋白質についてほぼ正確な空間上の距離を求めることができるという利点がある。

- 本発明の特徴は、このように各ステップで複数の計算方法を可能にしたことである。特に既存のシミュレーション手法による立体構造予測方法は時間がかかるという欠点を補うために、手法1のようにアミノ酸配列データから簡易的に各アミノ酸残基間の空間上の距離データを決定する方法を組み合わせることで、高速な計算を可能にして結合相手予測など大量のデータを処理する予測法を可能にしている。
- 15

- 次に、本発明は、当該距離データと各アミノ酸の電荷に従って蛋白質の全エネルギーを計算する(ステップSA3-2)。
- 20

- ここで、アミノ酸の電荷の決定方法も様々考えられる。例えば、正に帯電している荷電性アミノ酸(リジン、アルギニン)の電荷を1、負に帯電している荷電性アミノ酸(グルタミン酸、アスパラギン酸)の電荷を-1、それ以外のアミノ酸の電荷は0とするような方法がある。また、蛋白質構造データベースに登録された蛋白質の立体構造情報や、シミュレーション手法により得られた
- 25
- 立体構造情報を基にして既存の量子化学計算手法により各アミノ酸残基の電荷を決定することもできる。

また、蛋白質の全エネルギーの計算も様々な方法が考えられるが、例えば、  
 「Introduction to Computational Chemistry」(Frank Jensen 著 John Wiley & Sons 社 1999年)などで説明されている、分子力学、分子動力学、分子軌道法、密度汎関数法などのエネルギー計算法などの手法を用いても実施で  
 5 き、いずれかの手法から要求される予測精度と実施者の計算環境によって最適なものを選べば良い。他にもFragment MO法(Chemical Physics Letters, Volume 336, Issues 1-2, 9 March 2001, Pages 163-170)を用いるこ  
 10 とにより各アミノ酸残基のエネルギーを求めることができる。この方法は計算時間がかかるが、予測精度が高いことが期待される。

他に計算時間がかからない方法としては、以下のように静電エネルギーを計算する方法がある。

$$15 \quad E_{\text{total}} = 1/2 \sum \sum q_i q_j / r_{ij}$$

(i、j は全アミノ酸残基の任意のアミノ酸残基番号、i not j)

本式において、 $E_{\text{total}}$  は蛋白質の全エネルギーであり、 $q_i$  はアミノ酸残基 i の部分電荷であり、 $q_j$  はアミノ酸残基 j の部分電荷であり、 $r_{ij}$  はアミノ  
 20 酸残基 i とアミノ酸残基 j との空間上の距離である。

本式による方法は他の手法と比較して計算負荷が極めて少ないため、特に網羅的な計算の時に有効である。

次に、本発明は、特定のアミノ酸と蛋白質内のそれ以外のアミノ酸残基との相互作用エネルギーを以下の式により計算することにより、各アミノ酸残基が  
 25 蛋白質の全エネルギーをどれくらい安定化しているかを求める(ステップSA3-3)。

$$E_{\text{interaction}}(N) = q_N \sum q_j / r$$

$$E_{\text{total}} = 1/2 \sum E_{\text{interaction}}(N)$$

ここで、Nは任意のアミノ酸残基番号、 $E_{\text{interaction}}(N)$ はアミノ酸残基Nと  
 5 それ以外のアミノ酸残基との相互作用エネルギー、jはN以外のアミノ酸残基  
 番号、 $q_N$ はアミノ酸残基Nの部分電荷、 $q_j$ はアミノ酸残基jの部分電荷、r  
 はアミノ酸残基Nとアミノ酸残基jとの空間上の距離を示す。ここで全てのア  
 ミノ酸残基に対する相互作用エネルギーの総和の半分が蛋白質全体のエネルギ  
 ー $E_{\text{total}}$ に相当する。

10 次に、本発明は、ステップSA3-3で求めた相互作用エネルギーが相対的  
 に高いアミノ酸残基や、予め定めた閾値を超える相互作用エネルギーをもつア  
 ミノ酸残基を、エネルギー的に不安定なアミノ酸残基として特定することによ  
 り結合部位を予測する（ステップSA3-4）。

また、図52は、本発明により複数の蛋白質のアミノ酸配列情報からそれら  
 15 の蛋白質で複合体を形成するときの結合部位を予測する場合の概念を説明する  
 図である。

まず、本発明は、複数のアミノ酸配列上で結合部位となるアミノ酸残基（結  
 合残基）を仮定する（ステップSB3-1）。ここで、図57は、複数のアミ  
 ノ酸配列上で結合残基を仮定する場合の概念を示す図である。図57に示す例  
 20 においては、アミノ酸配列Aの50番目のアミノ酸残基と、アミノ酸配列Bの  
 100番目のアミノ酸残基とを結合残基と仮定している。ここで、結合残基は、  
 図51を用いて上述した本発明の手法により各アミノ酸配列において結合部位  
 であると予測されたアミノ酸残基を用いてもよい。

次に、本発明は、異なるアミノ酸配列上に存在する2つのアミノ酸残基間の  
 25 空間上の距離を決定する（ステップSB3-2）。ここで、空間上の距離の決  
 定手法には、上述した3つの手法を用いることができるが、以下に、最も計算

負荷が少なく効率的に計算を行うことができる 1) 高速計算手法を用いた場合について説明する。

まず、異なるアミノ酸配列上に存在する 2 つのアミノ酸残基間の配列上の距離を以下のように定義する。

5

(着目残基間の配列上の距離  $d$ ) =

$$\begin{aligned} & ( | \text{配列 A 上の着目残基と結合残基の配列上の距離} | \\ & + | \text{配列 B 上の着目残基と結合残基の配列上の距離} | ) \end{aligned}$$

10      ここで、図 5 8 は、着目残基の概念を説明する図である。図 5 8 に示すように、2 つのアミノ酸配列 (A および B) の結合残基と、結合残基以外の任意の着目残基を定義している。

ついで、本発明は、異なるアミノ酸配列上に存在する 2 つのアミノ酸残基間の配列上の距離  $d$  から複合体の立体構造の空間上の距離  $r$  を推定する (ステップ S B 3 - 3) 。

15

$$r = k d^n \quad (0 < n < 1)$$

ここで  $r$  は空間上の距離、 $d$  は配列上の距離、 $k$  は比例定数である。ここで、  
20       $n$  は 0 ~ 1 であるが、好ましくは、0.5 ~ 0.6 である。また、 $k$  は 2.8 Å ~ 4.8 Å であるが、好ましくは、3.3 Å ~ 4.3 Å である。つまり、配列上の距離  $d$  が分かれば、空間上の距離  $r$  を計算することができる。

この他に、複合体の立体構造が既知である場合には、上述した 2) 構造データ利用計算手法を用いて、アミノ酸残基間の空間上の距離を正確に求めること  
25      ができる。

また、上述した 3) シミュレーションデータ利用計算手法を用いることにより、複合体の立体構造を予測して、そのシミュレーションデータを利用するこ

とにより、アミノ酸残基間の空間上の距離をある程度正確に求めることができる。ここで、図62は、ドッキングシミュレーションを用いて複合体構造を生成する場合の概念を示す図である。図62に示すように、複数の立体構造情報を使って、複合体の構造を生成させるためにドッキングシミュレーションを行う。ドッキングシミュレーションは既知の様々なシミュレーション手法を用いることができる。例えば、図62で示すように、2つの蛋白質の距離と配向を変化させるものが一般的である。さらに具体的な例を挙げるとすると、一方の構造を固定し、もう一方の構造に回転で2つ、並進で2つの自由度を与えて、様々な構造を発生させる。2つの構造が重ならないが接触するという条件の構造を抽出すると、とりうる複合体の構造が得られる。

次に、本発明は、空間上の距離データと各アミノ酸の電荷に従って蛋白質の全エネルギーを計算する（ステップSB3-4）。

ここで、アミノ酸の電荷の決定方法も様々な考えられる。例えば、上述したように、正に帯電している荷電性アミノ酸（リジン、アルギニン）の電荷を1、負に帯電している荷電性アミノ酸（グルタミン酸、アスパラギン酸）の電荷を-1、それ以外のアミノ酸の電荷は0とするような方法がある。また、上述したように、蛋白質構造データベースに登録された複合体の立体構造情報や、シミュレーション手法により得られた複合体の立体構造情報を基にして既存の量子化学計算手法により各アミノ酸残基の電荷を決定することもできる。

また、蛋白質の全エネルギーの計算も様々な方法が考えられるが、上述したように、例えば、「Introduction to Computational Chemistry」（Frank Jensen著 John Wiley & Sons 社 1999年）などで説明されている、分子力学、分子動力学、分子軌道法、密度汎関数法などのエネルギー計算法などの手法を用いても実施でき、いずれかの手法から要求される予測精度と実施者の計算環境によって最適なものを選べば良い。他にも上述したように、Fragment MO法（Chemical Physics Letters, Volu

me 336, Issues 1-2, 9 March 2001, Pages 163-170) を用いることにより各アミノ酸残基のエネルギーを求めることができる。この方法は計算時間がかかるが、予測精度が高いことが期待される。

- 5 他に計算時間がかからない方法としては、上述したように、以下のように静電エネルギーを計算する方法がある。

$$E_{\text{total}} = \frac{1}{2} \sum_i \sum_j q_i q_j / r_{ij}$$

( $i, j$  は全アミノ酸残基の任意のアミノ酸残基番号、 $i \neq j$ )

10

本式において、 $E_{\text{total}}$  は蛋白質の全エネルギーであり、 $q_i$  はアミノ酸残基  $i$  の部分電荷であり、 $q_j$  はアミノ酸残基  $j$  の部分電荷であり、 $r_{ij}$  はアミノ酸残基  $i$  とアミノ酸残基  $j$  との空間上の距離である。このように本手法は基本的に二重線で示した処理フローと同様の手順で進み、候補蛋白質のアミノ酸配列を変えては繰り返す。その中で最も安定な複合体を形成できるものが相互作用相手の確率が高いと予測する。

15

次に、本発明は、ステップ S B 3-1 に戻り、相互作用するアミノ酸残基（結合残基）を変えて、全ての組合せについて  $E_{\text{total}}$  を計算して、最も  $E_{\text{total}}$  が低いときの結合残基を結合部位として予測する（ステップ S B 3-5）。

20 [システム構成]

まず、本システムの構成について説明する。図 5 3 は、本発明が適用される本システムの構成の一例を示すブロック図であり、該構成のうち本発明に関する部分のみを概念的に示している。本システムは、概略的に、結合部位予測装置 3100 と、配列情報等に関する外部データベースやホモロジー検索等の外部プログラム等を提供する外部システム 3200 とを、ネットワーク 3300 を介して通信可能に接続して構成されている。

25



図53においてネットワーク3300は、結合部位予測装置3100と外部システム3200とを相互に接続する機能を有し、例えば、インターネット等である。

図53において外部システム3200は、ネットワーク3300を介して、  
5 結合部位予測装置3100と相互に接続され、利用者に対してアミノ酸配列情報や蛋白質立体構造情報等に関する外部データベースやホモロジー検索やモチーフ検索等の外部プログラムを実行するウェブサイトを提供する機能を有する。

ここで、外部システム3200は、WEBサーバやASPサーバ等として構成してもよく、そのハードウェア構成は、一般に市販されるワークステーション、  
10 パーソナルコンピュータ等の情報処理装置およびその付属装置により構成してもよい。また、外部システム3200の各機能は、外部システム3200のハードウェア構成中のCPU、ディスク装置、メモリ装置、入力装置、出力装置、通信制御装置等およびそれらを制御するプログラム等により実現される。

図53において結合部位予測装置3100は、概略的に、結合部位予測装置  
15 3100の全体を統括的に制御するCPU等の制御部3102、通信回線等に接続されるルータ等の通信装置（図示せず）に接続される通信制御インターフェース部3104、入力装置3112や出力装置3114に接続される入出力制御インターフェース部3108、および、各種のデータベースやテーブルなどを格納する記憶部3106を備えて構成されており、これら各部は任意の通信路を介して通信可能に接続されている。さらに、この結合部位予測装置31  
20 00は、ルータ等の通信装置および専用線等の有線または無線の通信回線を介して、ネットワーク3300に通信可能に接続されている。

記憶部3106に格納される各種のデータベースやテーブル（アミノ酸配列データベース3106a～処理結果ファイル3106g）は、固定ディスク装置等のストレージ手段であり、各種処理に用いる各種のプログラムやテーブル  
25 やファイルやデータベースやウェブページ用ファイル等を格納する。

これら記憶部 3 1 0 6 の各構成要素のうち、アミノ酸配列データベース 3 1 0 6 a は、アミノ酸配列を格納したデータベースである。アミノ酸配列データベース 3 1 0 6 a は、インターネットを経由してアクセスする外部のアミノ酸配列データベースであってもよく、また、これらのデータベースをコピーしたり、オリジナルの配列情報を格納したり、さらに独自のアノテーション情報等を付加したりして作成したインハウスデータベースであってもよい。

また、蛋白質構造データベース 3 1 0 6 b は、蛋白質の立体構造情報を格納したデータベースである。蛋白質構造データベース 3 1 0 6 b は、インターネットを経由してアクセスする外部の立体構造情報データベースであってもよく、また、これらのデータベースをコピーしたり、オリジナルの立体構造情報を格納したり、さらに独自のアノテーション情報等を付加したりして作成したインハウスデータベースであってもよい。

また、距離データファイル 3 1 0 6 c は、アミノ酸配列に含まれる各アミノ酸残基間の距離（配列上の距離、空間上の距離）に関する情報等を格納する距離情報格納手段である。

また、全エネルギーデータファイル 3 1 0 6 d は、蛋白質の全エネルギーに関する情報等を格納する全エネルギーデータ格納手段である。

また、相互作用エネルギーデータファイル 3 1 0 6 e は、各アミノ酸残基の相互作用エネルギーに関する情報等を格納する相互作用エネルギーデータ格納手段である。

また、複合体構造データファイル 3 1 0 6 f は、複数の蛋白質の複合体構造に関する情報等を格納する複合体構造データ格納手段である。

また、処理結果ファイル 3 1 0 6 g は、結合部位予測装置 3 1 0 0 による各種の処理結果に関する情報等を格納する処理結果格納手段である。

また、図 5 3 において、通信制御インターフェース部 3 1 0 4 は、結合部位予測装置 3 1 0 0 とネットワーク 3 3 0 0（またはルータ等の通信装置）との

間における通信制御を行う。すなわち、通信制御インターフェース部 3 1 0 4 は、他の端末と通信回線を介してデータを通信する機能を有する。

また、図 5 3 において、入出力制御インターフェース部 3 1 0 8 は、入力装置 3 1 1 2 や出力装置 3 1 1 4 の制御を行う。ここで、出力装置 3 1 1 4 とし  
5 ては、モニタ（家庭用テレビを含む）の他、スピーカを用いることができる  
（なお、以下においては出力装置 3 1 1 4 をモニタとして記載する場合がある）。また、入力装置 3 1 1 2 としては、キーボード、マウス、および、マイク等を用いることができる。また、モニタも、マウスと協働してポインティングデバイス機能を実現する。

10 また、図 5 3 において、制御部 3 1 0 2 は、OS（O p e r a t i n g S y s t e m）等の制御プログラム、各種の処理手順等を規定したプログラム、および所要データを格納するための内部メモリを有し、これらのプログラム等により、種々の処理を実行するための情報処理を行う。制御部 3 1 0 2 は、機能概念的に、アミノ酸配列データ取得部 3 1 0 2 a、空間距離決定部 3 1 0 2  
15 b、電荷決定部 3 1 0 2 c、エネルギー計算部 3 1 0 2 d、候補アミノ酸残基決定部 3 1 0 2 e、複合体構造生成部 3 1 0 2 f、エネルギー最小化部 3 1 0 2 g、結合候補データ取得部 3 1 0 2 h、結合部位予測部 3 1 0 2 i、および、結合相手候補決定部 3 1 0 2 j を備えて構成されている。

このうち、アミノ酸配列データ取得部 3 1 0 2 a は、目的の蛋白質または生  
20 理活性ポリペプチドのアミノ酸配列データを取得するアミノ酸配列データ取得手段、目的の複数の蛋白質または生理活性ポリペプチドのアミノ酸配列データを取得するアミノ酸配列データ取得手段、および、目的の蛋白質または生理活性ポリペプチドのアミノ酸配列データと、結合候補となる複数の蛋白質または生理活性ポリペプチドのアミノ酸配列データとを取得するアミノ酸配列データ  
25 取得手段である。

また、空間距離決定部 3 1 0 2 b は、アミノ酸配列データ取得手段により取得されたアミノ酸配列データに含まれる各アミノ酸残基の間の空間上の距離を

決定する空間距離決定手段、アミノ酸配列データ取得手段により取得された複数のアミノ酸配列データに含まれる各アミノ酸残基の間の空間上の距離を、複合体構造生成手段により生成した複合体の立体構造情報に従って決定する空間距離決定手段、および、アミノ酸配列データ取得手段により取得された目的の

5 アミノ酸配列データと結合候補のアミノ酸配列データに含まれる各アミノ酸残基の間の空間上の距離を、複合体構造生成手段により生成した複合体の立体構造情報に従って決定する空間距離決定手段である。ここで、空間距離決定部 3 1 0 2 b は、図 5 4 に示すように、高速計算部 3 1 0 2 k、構造データ利用計算部 3 1 0 2 m、および、シミュレーションデータ利用計算部 3 1 0 2 n を備

10 えて構成されている。ここで、高速計算部 3 1 0 2 k は、高速計算手法により空間距離を決定する高速計算手段である。また、構造データ利用計算部 3 1 0 2 m は、構造データ利用計算手法により空間距離を決定する構造データ利用計算手段である。また、シミュレーションデータ利用計算部 3 1 0 2 n は、シミュレーションデータ利用計算手法により空間距離を決定するシミュレーション

15 データ利用計算手段である。

また、電荷決定部 3 1 0 2 c は、アミノ酸配列データに含まれる各アミノ酸残基が持つ電荷を決定する電荷決定手段、複数のアミノ酸配列データに含まれる各アミノ酸残基が持つ電荷を決定する電荷決定手段、および、目的のアミノ酸配列データと結合候補のアミノ酸配列データに含まれる各アミノ酸残基が持つ電荷を決定する電荷決定手段である。

20

また、エネルギー計算部 3 1 0 2 d は、空間距離決定手段により決定された各アミノ酸残基間の空間上の距離と、電荷決定手段により決定された各アミノ酸残基が持つ電荷とに従って、各アミノ酸残基のエネルギーを計算するエネルギー計算手段である。エネルギー計算部 3 1 0 2 d は、図 5 5 に示すように、

25 全エネルギー計算部 3 1 0 2 p、および、相互作用エネルギー計算部 3 1 0 2 q を備えて構成される。ここで、全エネルギー計算部 3 1 0 2 p は、蛋白質の全エネルギーを計算する全エネルギー計算手段である。また、相互作用エネル

ギー計算部 3 1 0 2 q は、アミノ酸残基の相互作用エネルギーを計算する相互作用エネルギー計算手段である。

また、候補アミノ酸残基決定部 3 1 0 2 e は、エネルギー計算手段により計算されたエネルギーに従って、結合部位となる候補アミノ酸残基を決定する候補アミノ酸残基決定手段、および、エネルギー最小化手段によりエネルギーの総和が最小となる結合部位を、結合部位の候補アミノ酸残基として決定する候補アミノ酸残基決定手段である。

また、複合体構造生成部 3 1 0 2 f は、目的の複数の蛋白質または生理活性ポリペプチドの結合した複合体の立体構造情報を生成する複合体構造生成手段、および、目的の蛋白質または生理活性ポリペプチドと、結合候補となる蛋白質または生理活性ポリペプチドの結合した複合体の立体構造情報を生成する複合体構造生成手段である。

また、エネルギー最小化部 3 1 0 2 g は、複合体構造生成手段により複合体について結合部位を変えて複合体の立体構造情報を生成し、エネルギー計算手段により各アミノ酸残基のエネルギーを計算し、当該エネルギーの総和が最小となる結合部位を求めるエネルギー最小化手段である。

また、結合候補データ取得部 3 1 0 2 h は、結合候補となる蛋白質のアミノ酸配列データ等を取得する結合候補データ取得手段である。

また、結合部位予測部 3 1 0 2 i は、結合部位の候補アミノ酸残基から結合部位のアミノ酸残基を予測する結合部位予測手段である。

また、結合相手候補決定部 3 1 0 2 j は、全ての結合候補についてエネルギー最小化手段を実行した結果、エネルギーの総和が最小となる結合部位をもつ結合候補を決定する結合候補決定手段である。

これら各部によって行なわれる処理の詳細については、後述する。

[システムの処理]

次に、このように構成された本実施の形態における本システムの処理の一例について、以下に図 5 3 ～図 7 1 を参照して詳細に説明する。

図59は、本実施形態における本システムの処理の一例を示すフローチャートである。図59において、点線で示す処理フローは本システムにより一つの蛋白質の配列中における結合部位を予測する場合の処理フローを示し、また、二重線で示す処理フローは本システムにより相互作用することが既知である複数の蛋白質のアミノ酸配列を用いて結合部位を予測する場合の処理フローを示し、さらに、実線で示す処理フローは本システムにより目的蛋白質に対して最も良く結合する相手側の候補蛋白質を予測する場合の処理フローを示す。これらの3つの処理フローの基本的な考え方、計算の処理のほとんどが共通である。さらには、相互作用情報を解析するという大目的については共通である。

10 [一つの蛋白質の配列中における結合部位を予測する場合の処理]

次に、本システムにより一つの蛋白質の配列中における結合部位を予測する場合の処理の詳細について図59等を参照して説明する。図59の点線で示す処理フローは、本実施形態における本システムにより一つの蛋白質の配列中における結合部位を予測する場合の処理の一例を示すフローチャートである。

15 まず、結合部位予測装置3100は、アミノ酸配列データ取得部3102aの処理により、Genbankなどの外部システム3200の外部データベースやアミノ酸配列データベース3106aなどにアクセスして、目的の蛋白質または生理活性ポリペプチドのアミノ酸配列データを取得する(ステップSC3-1)。

20 そして、結合部位予測装置3100は、空間距離決定部3102bの処理により、ステップSC3-1により取得されたアミノ酸配列データに含まれる各アミノ酸残基の間の空間上の距離を決定する(ステップSC3-2)。

ここで、空間距離決定部3102bは、高速計算部3102kの処理により上述した高速計算手法により各アミノ酸残基の間の配列上の距離から空間上の距離を決定してもよく、また、構造データ利用計算部3102mの処理により  
25 上述した構造データ利用計算手法により既知の構造データを用いて各アミノ酸残基の間の空間上の距離を決定してもよく、さらに、シミュレーションデータ

利用計算部 3 1 0 2 n の処理により上述したシミュレーションデータ利用計算手法により既存の構造シミュレーションプログラムの処理結果による予測構造を用いて各アミノ酸残基の間の空間上の距離を決定してもよい。

5 ついで、結合部位予測装置 3 1 0 0 は、電荷決定部 3 1 0 2 c の処理により、アミノ酸配列データに含まれる各アミノ酸残基が持つ電荷を決定する（ステップ S C 3 - 3）。アミノ酸の電荷の決定方法も様々考えられる。一般的には、正に帯電している荷電性アミノ酸（リジン、アルギニン）を 1、負に帯電している荷電性アミノ酸（グルタミン酸、アスパラギン酸）を -1、それ以外は 0 とするような方法がある。また、得られた立体構造情報を基にして、量子化学  
10 計算で電荷を決定することもできる。さらに、実験的に各アミノ酸残基の電荷に関する実験データが分かっている場合は、それを反映することが好ましい。

ついで、結合部位予測装置 3 1 0 0 は、エネルギー計算部 3 1 0 2 d の処理により、決定された各アミノ酸残基間の空間上の距離と、各アミノ酸残基が持つ電荷とに従って、各アミノ酸残基のエネルギーを計算する（ステップ S C 3  
15 - 4）。

ここで、エネルギー計算には様々な手法があるが、最も計算時間がかからない方法である静電エネルギーを計算する方法について以下に説明する。

まず、全エネルギー計算部 3 1 0 2 p の処理により、以下の式により蛋白質の全エネルギーを計算する。

20

$$E_{total} = 1/2 \sum \sum q_i q_j / r_{ij}$$

（i、j は全アミノ酸残基の任意のアミノ酸残基番号、i not j）

本式において、 $E_{total}$  は蛋白質の全エネルギーであり、 $q_i$  はアミノ酸残基  
25 i の部分電荷であり、 $q_j$  はアミノ酸残基 j の部分電荷であり、 $r_{ij}$  はアミノ酸残基 i とアミノ酸残基 j との空間上の距離である。

次に、相互作用エネルギー計算部 3102q は、特定のアミノ酸と蛋白質内のそれ以外のアミノ酸残基との相互作用エネルギーを以下の式により計算することにより、各アミノ酸残基が蛋白質の全エネルギーをどれくらい安定化しているかを求める。

5

$$E_{\text{interaction}}(N) = q_N \sum q_j / r$$

$$E_{\text{total}} = 1/2 \sum E_{\text{interaction}}(N)$$

ここで、Nは任意のアミノ酸残基番号、 $E_{\text{interaction}}(N)$ はアミノ酸残基Nと  
 10 それ以外のアミノ酸残基との相互作用エネルギー、jはN以外のアミノ酸残基番号、 $q_N$ はアミノ酸残基Nの部分電荷、 $q_j$ はアミノ酸残基jの部分電荷、rはアミノ酸残基Nとアミノ酸残基jとの空間上の距離を示す。ここで全てのアミノ酸残基に対する相互作用エネルギーの総和の半分が蛋白質全体のエネルギー $E_{\text{total}}$ に相当する。

15 そして、結合部位予測装置 3100 は、候補アミノ酸残基決定部 3102e の処理により、計算された相互作用エネルギーに従って、結合部位となる候補アミノ酸残基を決定する（ステップSC3-5）。すなわち、候補アミノ酸残基決定部 3102e は、相互作用エネルギーが相対的に高いアミノ酸残基や、予め定めた閾値を超える相互作用エネルギーをもつアミノ酸残基を、エネルギー  
 20 的に不安定なアミノ酸残基として特定することにより、結合部位となる候補アミノ酸残基を決定する。

そして、結合部位予測装置 3100 は、結合部位予測部 3102i の処理により、候補アミノ酸残基のうち空間的、または、エネルギー的に結合部位とならない候補を除去することにより、結合部位を予測する。例えば、処理結果の  
 25 候補アミノ酸残基のエネルギーなどの一例として図60のような結果が得られるとすると、結合部位予測部 3102i は、図60の中で最もエネルギーの高いグルタミン酸（GLU）を一番の結合部位の候補として予測する。また、結



合部位予測部 3102 i は、例えば図 6 1 に示すように立体構造上で不安定な部分がクラスター化しているもの（黒で示したアミノ酸残基部分）が結合部位である可能性が高いと予測する。

これにて、本システムにより一つの蛋白質の配列中における結合部位を予測  
5 する場合の処理が終了する。

[相互作用することが既知である複数の蛋白質のアミノ酸配列を用いて結合部位を予測する場合の処理]

次に、本システムにより相互作用することが既知である複数の蛋白質のアミノ酸配列を用いて結合部位を予測する場合の処理の詳細について図 5 9 等を参照して説明する。図 5 9 の二重線で示す処理フローは、本実施形態における本  
10 システムにより相互作用することが既知である複数の蛋白質のアミノ酸配列を用いて結合部位を予測する場合の処理の一例を示すフローチャートである。

まず、結合部位予測装置 3100 は、アミノ酸配列データ取得部 3102 a の処理により、G e n b a n k などの外部システム 3200 の外部データベース  
15 スやアミノ酸配列データベース 3106 a などにアクセスして、目的の蛋白質または生理活性ポリペプチドのアミノ酸配列データを取得する（ステップ S C 3-1）。

ついで、結合部位予測装置 3100 は、複合体構造生成部 3102 f の処理により、目的の複数の蛋白質または生理活性ポリペプチドの結合した複合体の  
20 立体構造情報を生成する（ステップ S C 3-7）。ここで、図 6 2 を用いて上述したように、複合体構造生成部 3102 f は、シミュレーションデータ利用計算手法を用いることにより、複合体の立体構造を予測してもよい。また、複合体構造生成部 3102 f は、複合体の立体構造が既知である場合には、複合体の立体構造情報を取得してもよい。

25 また、複合体構造生成部 3102 f は、上述したように、複数のアミノ酸配列上で結合部位となるアミノ酸残基（結合残基）を仮定することにより、実際に複合体構造を生成することなく、処理を進めることができる。ここで、図 5

7 は、複数のアミノ酸配列上で結合残基を仮定する場合の概念を示す図である。  
 図 5 7 に示す例においては、アミノ酸配列 A の 5 0 番目のアミノ酸残基と、ア  
 ミノ酸配列 B の結合部位予測装置 1 0 0 番目のアミノ酸残基とを結合残基と仮  
 定している。ここで、結合残基は、上述した本発明の手法により各アミノ酸配  
 5 列において結合部位であると予測されたアミノ酸残基を用いてもよい。

ついで、結合部位予測装置 3 1 0 0 は、空間距離決定部 3 1 0 2 b の処理に  
 より、取得された複数のアミノ酸配列データに含まれる各アミノ酸残基の間の  
 空間上の距離を、複合体の立体構造情報に従って決定する（ステップ S C 3 -  
 2）。

10 ここで、空間上の距離の決定手法には、上述した 3 つの手法を用いることが  
 できるが、複合体の立体構造が既知である場合や、ドッキングシミュレーショ  
 ンを行った際には、空間距離決定部 3 1 0 2 b は、アミノ酸残基間の空間上の  
 距離を正確に求めることができる。以下に、最も計算負荷が少なく効率的に計  
 算を行うことができる 1) 高速計算手法を用いた場合について説明する。

15 まず、空間距離決定部 3 1 0 2 b は、異なるアミノ酸配列上に存在する 2 つ  
 のアミノ酸残基間の配列上の距離を以下のように定義する。

（着目残基間の配列上の距離 d）＝

20 （ | 配列 A 上の着目残基と結合残基の配列上の距離 |  
 + | 配列 B 上の着目残基と結合残基の配列上の距離 | ）

ここで、図 5 8 は、着目残基の概念を説明する図である。図 5 8 に示すよう  
 に、2 つのアミノ酸配列（A および B）の結合残基と、結合残基以外の任意の  
 着目残基を定義している。

25 ついで、空間距離決定部 3 1 0 2 b は、異なるアミノ酸配列上に存在する 2  
 つのアミノ酸残基間の配列上の距離 d から複合体の立体構造の空間上の距離 r  
 を推定する。

$$r = k d^n \quad (0 < n < 1)$$

ここで  $r$  は空間上の距離、 $d$  は配列上の距離、 $k$  は比例定数である。ここで、  
 5  $n$  は  $0 \sim 1$  であるが、好ましくは、 $0.5 \sim 0.6$  である。また、 $k$  は  $2.8 \text{ \AA} \sim 4.8 \text{ \AA}$  であるが、好ましくは、 $3.3 \text{ \AA} \sim 4.3 \text{ \AA}$  である。

ついで、結合部位予測装置 3100 は、電荷決定部 3102 c の処理により、  
 複数のアミノ酸配列データに含まれる各アミノ酸残基が持つ電荷を決定する  
 （ステップ SC3-3）。

10 ついで、結合部位予測装置 3100 は、エネルギー計算部 3102 d の処理  
 により、ステップ SC3-2 により決定された各アミノ酸残基間の空間上の距  
 離と、ステップ SC3-3 により決定された各アミノ酸残基が持つ電荷とに従  
 って、各アミノ酸残基のエネルギーを計算する（ステップ SC3-4）。

そして、結合部位予測装置 3100 は、候補アミノ酸残基決定部 3102 e  
 15 の処理により、計算された相互作用エネルギーに従って、結合部位となる候補  
 アミノ酸残基を決定する（ステップ SC3-5）。

そして、結合部位予測装置 3100 は、エネルギー最小化部 3102 g の処  
 理により、ステップ SC3-7 により複合体について結合部位を変えて複合体  
 の立体構造情報を生成し、ステップ SC3-4 により各アミノ酸残基のエネル  
 20 ギーを計算し、当該エネルギーの総和が最小となる結合部位を求める（ステッ  
 プ SC3-7 からステップ SC3-5 までの各ステップを適宜繰り返す）。

そして、結合部位予測装置 3100 は、候補アミノ酸残基決定部 3102 e  
 の処理により、最終的にエネルギーの総和が最小となる結合部位を、結合部位  
 の候補アミノ酸残基として決定する（ステップ SC3-5）。ここで、候補ア  
 25 ミノ酸残基決定部 3102 e は、蛋白質のエネルギーの総和をアミノ酸配列に  
 よりプロットした図を作成して出力装置 3114 に出力してもよい。図 63 は、  
 蛋白質 A と蛋白質 B の各アミノ酸残基を結合残基とした場合のエネルギーの総

和をプロットした図の一例である。このようにプロット図を作成することにより、2つのアミノ酸配列のどのアミノ酸残基を結合残基として選択するとエネルギーの総和が少なくなるかを視覚的に把握することができるようになる。

これにて、本システムにより相互作用することが既知である複数の蛋白質の  
5 アミノ酸配列を用いて結合部位を予測する場合の処理が終了する。

[目的蛋白質に対して最も良く結合する相手側の候補蛋白質を予測する場合の処理]

次に、本システムにより目的蛋白質に対して最も良く結合する相手側の候補蛋白質を予測する場合の処理の詳細について図59等を参照して説明する。図  
10 59の実線で示す処理フローは、本実施形態における本システムにより目的蛋白質に対して最も良く結合する相手側の候補蛋白質を予測する場合の処理の一例を示すフローチャートである。

まず、結合部位予測装置3100は、アミノ酸配列データ取得部3102aの処理により、Genbankなどの外部システム3200の外部データベースやアミノ酸配列データベース3106aなどにアクセスして、目的の蛋白質  
15 または生理活性ポリペプチドのアミノ酸配列データを取得する(ステップSC3-1)。また、結合部位予測装置3100は、結合候補データ取得部3102hの処理により、Genbankなどの外部システム3200の外部データベースやアミノ酸配列データベース3106aなどにアクセスして、目的の蛋白質の結合候補となる1つまたは複数の蛋白質または生理活性ポリペプチドの  
20 アミノ酸配列データを取得する(ステップSC3-6)。

ついで、結合部位予測装置3100は、複合体構造生成部3102fの処理により、目的の蛋白質または生理活性ポリペプチドと、結合候補となる蛋白質または生理活性ポリペプチドとが結合した複合体の立体構造情報を生成する  
25 (ステップSC3-7)。

ついで、結合部位予測装置3100は、空間距離決定部3102bの処理により、ステップSC3-1より取得された目的のアミノ酸配列データと、ステ

ップSC3-6により取得された結合候補のアミノ酸配列データとに含まれる各アミノ酸残基の間の空間上の距離を、ステップSC3-7により生成した複合体の立体構造情報に従って決定する（ステップSC3-2）。

5 ついで、結合部位予測装置3100は、電荷決定部3102cの処理により、目的のアミノ酸配列データと結合候補のアミノ酸配列データに含まれる各アミノ酸残基が持つ電荷を決定する（ステップSC3-3）。

10 ついで、結合部位予測装置3100は、エネルギー計算部3102dの処理により、ステップSC3-2により決定された各アミノ酸残基間の空間上の距離と、ステップSC3-3より決定された各アミノ酸残基が持つ電荷とに従って、各アミノ酸残基のエネルギーを計算する（ステップSC3-4）。

15 ついで、結合部位予測装置3100は、エネルギー最小化部3102gの処理により、ステップSC3-7により複合体について結合部位を変えて複合体の立体構造情報を生成し、ステップSC3-4により各アミノ酸残基のエネルギーを計算し、当該エネルギーの総和が最小となる結合部位を求める（ステップSC3-7からステップSC3-5までを適宜繰り返す）。

20 ついで、結合部位予測装置3100は、候補アミノ酸残基決定部3102eの処理により、全ての結合候補についてステップSC3-6からステップSC3-5を繰り返して、エネルギー最小化処理を実行した結果、エネルギーの総和が最小となる結合部位をもつ結合候補を決定する（ステップSC3-8）。

これにて、本システムにより目的蛋白質に対して最も良く結合する相手側の候補蛋白質を予測する場合の処理が終了する。

#### [本発明の実施例]

次に本発明の実施例の詳細を図64～図71を参照して説明する。

#### [本発明の実施例1；リボヌクレアーゼA]

25 次に、本発明の実施例1の詳細について図64～図66等を参照して説明する。実施例1は単体の蛋白質の結合部位予測に関する。

加水分解酵素であるリボヌクレアーゼAは実験的に良く調べられている蛋白質である。リボヌクレアーゼAは、そのインヒビターとの複合体の構造が分かっているので、アミノ酸配列上で結合部位が特定されている。

まず、リボヌクレアーゼAのアミノ酸配列データを蛋白質シーケンスデータベース Genbank から取得した。

そして、リボヌクレアーゼAのアミノ酸配列データからアミノ酸の距離情報を以下の方法により推測した。まず、PDB (Protein Data Bank) に登録されている全ての蛋白質、または、ポリペプチドの立体構造情報から、アミノ酸の種類ごとに配列上の距離と空間上の距離との関連性を求めた。例えば、図64は2つのグルタミン酸の配列上の距離と空間上の距離との関係を示す図である。図64に示すように、例えば、グルタミン酸とグルタミン酸が配列上20残基離れていた時の空間距離の平均は20 Åであるということが既知の統計学的手法により分かる。以上のようにして、アミノ酸残基間の配列上の距離と空間上の距離との関連を示す情報を得た。

そして、アミノ酸の電荷を決定した。ここではグルタミン酸、アスパラギン酸を-1、アルギニン、リジン、ヒスチジンを+1、その他は0というようにアミノ酸残基ごとに電荷を割り当てた。

そして、各アミノ酸残基の相互作用エネルギーを以下のように計算した。

$$E_{\text{interaction}}(K) = q_K \sum q_j / r$$

(ここで、Kはアミノ酸残基番号、 $E_{\text{interaction}}(K)$ はアミノ酸残基Kとそれ以外との相互作用エネルギー、jはK以外の任意のアミノ酸残基、rはアミノ酸残基Kとアミノ酸残基jとの空間上の距離)

そして、上式によってリボヌクレアーゼAのアミノ酸残基ごとのエネルギーを計算し、リボヌクレアーゼAの各アミノ酸残基のエネルギーをアミノ酸残基

番号に対してプロットした。図 6 5 は、リボヌクレアーゼ A の各アミノ酸残基のエネルギーをアミノ酸残基番号に対してプロットした図である。

そして、リボヌクレアーゼ A のアミノ酸残基のエネルギーが 0 以上のものを結合部位候補として表にまとめた (図 6 6) 。図 6 6 に示すように結合部位候補 1 8 個のうち、実際の結合部位 (実験的に求められた結合部位) は 1 2 個だった。このように、本発明により、リボヌクレアーゼ A のアミノ酸配列情報だけを  
5 用いて、非常に高精度で、かつ、高速に結合部位を予測することができた。

これにて、本発明の実施例 1 が終了する。

[本発明の実施例 2 ; アセチルコリン・エステラーゼ・インヒビター]

10 次に、本発明の実施例 2 の詳細について図 6 7 ~ 図 6 9 等を参照して説明する。実施例 2 も単体の蛋白質の結合部位予測に関する。

実施例 2 では、アセチルコリン・エステラーゼ・インヒビターのアミノ酸配列を基にして、結合部位を推定した。ここでは立体構造予測をするのではなく、PDB に含まれる既存の立体構造情報データを利用した。

15 図 6 7 は、PDB に格納されたアセチルコリン・エステラーゼ・インヒビターの立体構造情報データの一部を示す図である。図 6 7 の 2 列目からそれぞれ原子番号、原子種、chain 名、アミノ酸残基番号、X 座標、Y 座標、Z 座標を示している。

例えばアミノ酸残基番号 I とアミノ酸残基番号 J との空間上の距離  $R_{IJ}$  は、  
20 アミノ酸残基番号 I の重心や特定の主鎖の原子などの座標を  $(x_I, y_I, z_I)$  とし、アミノ酸残基番号 J の重心や特定の主鎖の原子などの座標を  $(x_J, y_J, z_J)$  とすると、以下のように計算することができる。

$$R_{IJ}^2 = (x_I - x_J)^2 + (y_I - y_J)^2 + (z_I - z_J)^2$$

25 (ここで、 $R_{IJ} > 0$ )

具体的に説明すると、図 6 7 に示すアミノ酸残基番号 4 のグルタミン酸と、アミノ酸残基番号 5 のアスパラギン酸との空間上の距離を  $\alpha$  カーボンの原子間の距離によって計算すると、以下のようになる。

$$\begin{aligned}
 5 \quad R_{45}^2 &= (32.664 - 36.279)^2 \\
 &+ (8.451 - 7.196)^2 \\
 &+ (205.542 - 205.808)^2 \\
 &= 14.714 \\
 R_{45} &= 3.835884
 \end{aligned}$$

10

また、図 6 8 は本発明により求めたアセチルコリン・エステラーゼ・インヒビターのエネルギーを示す図である。図 6 8 からアセチルコリン・エステラーゼ・インヒビターのエネルギーが 0 以上のもの 10 個を結合部位候補として取り出し、実際に結合部位かどうかを実験的に調べると、その内に 7 個は実際に

15

結合部位であった（図 6 9）。

以上のように極めて高精度に結合部位を予測することができる。実施例 2 では既知の立体構造情報を利用した点が実施例 1 とは異なる点である。つまり、実施例 1 と実施例 2 では空間距離の決定手法を変えているが、いずれも良好な結果を与えたことから、どのような空間距離の決定手法を使っても、本発明の

20

これにて、本発明の実施例 2 が終了する。

[本発明の実施例 3 ; 「huntingtin-associated protein interacting protein」と、「nitric oxide synthase 2A」との複合体]

25

次に、本発明の実施例 3 の詳細について図 7 0 等を参照して説明する。実施例 3 は 2 つの蛋白質が結合する時の結合部位予測に関する。「huntingtin-associated protein interacting



protein」は、「nitric oxide synthase 2 A」と結合することが実験的に分かっている。更に、結合部位は「huntingtin-associated protein interacting protein」はアミノ酸残基番号600付近であり、「nitric oxide synthase 2 A」はアミノ酸残基番号100付近であることが分かっている。

ここで、本実施例においても実施例1と同様にして配列情報の獲得、立体構造の予測、電荷の決定を行った。ただし、アミノ酸間の配列上の距離から空間的な距離に変換する方法は、蛋白質がガウス鎖をとると仮定し、配列上の距離と空間上の距離とを以下の式で関係づけた。

$$r = 3.8 d^{0.5}$$

ここでrは空間上の距離、dは配列上の距離である。

また、複合体構造生成は上述した高速計算手法を使った。つまり、次式を用いた。

$$(\text{空間上の距離}) = k \left( |A \text{ 上の着目残基と結合残基の配列上の距離}| + |B \text{ 上の着目残基と結合残基の配列上の距離}| \right)^n$$

20

そして、それぞれの結合部位を仮定した複合体のエネルギーを計算し、図70を作成することができた。ここで、図70は横軸に、huntingtin-associated protein interacting proteinの結合部位のアミノ酸残基番号をとり、縦軸にnitric oxide synthase 2 Aの結合部位のアミノ酸残基番号をとり、各結合部位で複合体を生成したときのエネルギーの総和を等高線表示したものである。

図70によれば、例えば、huntingtin-associated protein interacting proteinのアミノ酸残基500番と、nitric oxide synthase 2Aのアミノ酸残基150番とが結合部位の場合の複合体のエネルギーは-10というように、  
5 結合部位ごとのエネルギーが求められる。

図70に示すように、エネルギーの極小部は2つあり、1つはhuntingtin-associated protein interacting proteinの結合部位がアミノ酸残基600～950付近で、nitric oxide synthase 2Aの結合部位がアミノ酸残基25  
10 ～100付近で結合した場合であり、もう1つはhuntingtin-associated protein interacting proteinの結合部位がアミノ酸残基650～900付近で、nitric oxide synthase 2Aの結合部位がアミノ酸残基475～500付近で結合した場合である。

15 ここで、前者が実際の結合部位に対応している（図中の黒丸で囲んだ部分）。  
以上のように、2つの蛋白質の結合部位を正確に予測することができた。

これにて、本発明の実施例3が終了する。

[本発明の実施例4；E2F transcription factor  
1]

20 次に、本発明の実施例4の詳細について図71等を参照して説明する。

実施例4は結合相手予測に関する。E2F transcription factor 1（以下、E2F1）は相互作用相手に関する情報が実験的に良く調べられている蛋白質である。

ここで、E2F1と相互作用する相手をHomo Sapiensの遺伝子のデータベース（ここでは無作為に6600個を抽出した）から探し、候補蛋白質のアミノ酸配列データとした。  
25

そして、実施例 3 で計算した手順と同様に、相手の候補蛋白質ごとに E 2 F 1 との結合部位を探す。そして、最もエネルギーが安定（最小）になった時のエネルギーを相互作用エネルギーとする。図 7 1 は各候補蛋白質の相互作用エネルギーと遺伝子の数のヒストグラムである。

5 図 7 1 に示したように、相対的な相互作用エネルギーを計算することができた。例えば、相互作用エネルギーが 9 0 よりも大きい（エネルギーが - 9 0 よりも小さい）ものは 1 0 0 個あるが、ここにあるものが相互作用相手である可能性が高い。この方法は非常に高速で、相互作用を網羅的に計算することができる。

10 これにて、本発明の実施例 4 が終了する。

[他の実施の形態]

さて、これまで本発明の実施の形態について説明したが、本発明は、上述した実施の形態以外にも、上記特許請求の範囲に記載した技術的思想の範囲内において種々の異なる実施の形態にて実施されてよいものである。

15 例えば、結合部位予測装置 3 1 0 0 がスタンドアローンの形態で処理を行う場合を一例に説明したが、結合部位予測装置 3 1 0 0 とは別筐体で構成されるクライアント端末からの要求に応じて処理を行い、その処理結果を当該クライアント端末に返却するように構成してもよい。

また、実施形態において説明した各処理のうち、自動的に行なわれるものとして説明した処理の全部または一部を手動的に行うこともでき、あるいは、手動的に行なわれるものとして説明した処理の全部または一部を公知の方法で自動的に行うこともできる。

この他、上記文書中や図面中で示した処理手順、制御手順、具体的名称、各種の登録データや検索条件等のパラメータを含む情報、画面例、データベース構成については、特記する場合を除いて任意に変更することができる。

25 また、結合部位予測装置 3 1 0 0 に関して、図示の各構成要素は機能概念的なものであり、必ずしも物理的に図示の如く構成されていることを要しない。

例えば、結合部位予測装置 3 1 0 0 の各部または各装置が備える処理機能、特に制御部 3 1 0 2 にて行なわれる各処理機能については、その全部または任意の一部を、CPU (Central Processing Unit) および当該 CPU にて解釈実行されるプログラムにて実現することができ、ある  
5 いは、ワイヤードロジックによるハードウェアとして実現することも可能である。なお、プログラムは、後述する記録媒体に記録されており、必要に応じて結合部位予測装置 3 1 0 0 に機械的に読み取られる。

すなわち、ROM または HD などの記憶部 3 1 0 6 などには、OS (Operating System) と協働して CPU に命令を与え、各種処理を行うためのコンピュータプログラムが記録されている。このコンピュータプログラムは、RAM 等にロードされることによって実行され、CPU と協働して制御部 3 1 0 2 を構成する。また、このコンピュータプログラムは、結合部位予測装置 3 1 0 0 に対して任意のネットワーク 3 3 0 0 を介して接続されたアプリケーションプログラムサーバに記録されてもよく、必要に応じてその全部または一部をダウンロードすることも可能である。  
10  
15

また、本発明にかかるプログラムを、コンピュータ読み取り可能な記録媒体に格納することもできる。ここで、この「記録媒体」とは、フレキシブルディスク、光磁気ディスク、ROM、EPROM、EEPROM、CD-ROM、MO、DVD 等の任意の「可搬用の物理媒体」や、各種コンピュータシステム  
20 に内蔵される ROM、RAM、HD 等の任意の「固定用の物理媒体」、あるいは、LAN、WAN、インターネットに代表されるネットワークを介してプログラムを送信する場合の通信回線や搬送波のように、短期にプログラムを保持する「通信媒体」を含むものとする。

また、「プログラム」とは、任意の言語や記述方法にて記述されたデータ処理方法であり、ソースコードやバイナリコード等の形式を問わない。なお、  
25

「プログラム」は必ずしも単一的に構成されるものに限られず、複数のモジュールやライブラリとして分散構成されるものや、OS (Operating

S y s t e m) に代表される別個のプログラムと協働してその機能を達成するものをも含む。なお、実施の形態に示した各装置において記録媒体を読み取るための具体的な構成、読み取り手順、あるいは、読み取り後のインストール手順等については、周知の構成や手順を用いることができる。

- 5        記憶部 3 1 0 6 に格納される各種のデータベース等（アミノ酸配列データベース 3 1 0 6 a ～処理結果ファイル 3 1 0 6 g）は、RAM、ROM等のメモリ装置、ハードディスク等の固定ディスク装置、フレキシブルディスク、光ディスク等のストレージ手段であり、各種処理やウェブサイト提供に用いる各種のプログラムやテーブルやファイルやデータベースやウェブページ用ファイル  
10        等を格納する。

- また、結合部位予測装置 3 1 0 0 は、既知のパーソナルコンピュータ、ワークステーション等の情報処理端末等の情報処理装置にプリンタやモニタやイメージスキャナ等の周辺装置を接続し、該情報処理装置に本発明の方法を実現させるソフトウェア（プログラム、データ等を含む）を実装することにより実現  
15        してもよい。

- さらに、結合部位予測装置 3 1 0 0 の分散・統合の具体的形態は図示のものに限られず、その全部または一部を、各種の負荷等に応じた任意の単位で、機能的または物理的に分散・統合して構成することができる。例えば、各データベースを独立したデータベース装置として独立に構成してもよく、また、処理  
20        の一部を C G I（C o m m o n   G a t e w a y   I n t e r f a c e）を用いて実現してもよい。

- また、ネットワーク 3 3 0 0 は、結合部位予測装置 3 1 0 0 と外部システム 3 2 0 0 とを相互に接続する機能を有し、例えば、インターネットや、イントラネットや、LAN（有線／無線の双方を含む）や、VANや、パソコン通信  
25        網や、公衆電話網（アナログ／デジタルの双方を含む）や、専用回線網（アナログ／デジタルの双方を含む）や、CATV網や、IMT 2 0 0 0 方式、GSM方式またはPDC／PDC－P方式等の携帯回線交換網／携帯パケット交換

網や、無線呼出網や、Bluetooth等の局所無線網や、PHS網や、CS、BSまたはISDB等の衛星通信網等のうちいずれかを含んでもよい。すなわち、本システムは、有線・無線を問わず任意のネットワークを介して、各種データを送受信することができる。

- 5      以上詳細に説明したように、本発明によれば、蛋白質または生理活性ポリペプチドのアミノ酸配列データから当該蛋白質または生理活性ポリペプチドの立体構造における各アミノ酸残基間の空間上の距離データを求め、当該距離データと各アミノ酸の電荷に従って静電的に不安定なアミノ酸残基を特定することにより結合部位を予測するので、蛋白質または生理活性ポリペプチドのアミノ
- 10      酸配列から静電的に不安定になりそうなアミノ酸残基が結合部位になりやすいことを利用して、高速かつ高精度に結合部位を予測することができる結合部位予測装置、結合部位予測方法、プログラム、および、記録媒体を提供することができる。

- 15      また、本発明によれば、目的の蛋白質または生理活性ポリペプチドのアミノ酸配列データを取得し、取得されたアミノ酸配列データに含まれる各アミノ酸残基の間の空間上の距離を決定し、アミノ酸配列データに含まれる各アミノ酸残基が持つ電荷を決定し、決定された各アミノ酸残基間の空間上の距離と、決定された各アミノ酸残基が持つ電荷とに従って、各アミノ酸残基のエネルギーを計算し、計算されたエネルギーに従って、結合部位となる候補アミノ酸残基
- 20      を決定するので、蛋白質または生理活性ポリペプチドのアミノ酸配列から静電的に不安定になりそうなアミノ酸残基が結合部位になりやすいことを利用して、高速かつ高精度に結合部位を予測することができる結合部位予測装置、結合部位予測方法、プログラム、および、記録媒体を提供することができる。

- 25      また、本発明によれば、目的の複数の蛋白質または生理活性ポリペプチドのアミノ酸配列データを取得し、目的の複数の蛋白質または生理活性ポリペプチドの結合した複合体の立体構造情報を生成し、取得された複数のアミノ酸配列データに含まれる各アミノ酸残基の間の空間上の距離を、生成した複合体の立

体構造情報に従って決定し、複数のアミノ酸配列データに含まれる各アミノ酸残基が持つ電荷を決定し、決定された各アミノ酸残基間の空間上の距離と、決定された各アミノ酸残基が持つ電荷とに従って、各アミノ酸残基のエネルギーを計算し、複合体について結合部位を変えて複合体の立体構造情報を生成し、  
5 各アミノ酸残基のエネルギーを計算し、当該エネルギーの総和が最小となる結合部位を求め、エネルギーの総和が最小となる結合部位を、結合部位の候補アミノ酸残基として決定するので、蛋白質または生理活性ポリペプチドのアミノ酸配列から静電的に不安定になりそうなアミノ酸残基が結合部位になりやすいことを利用して、高速かつ高精度に結合部位を予測することができる結合部位  
10 予測装置、結合部位予測方法、プログラム、および、記録媒体を提供することができる。

さらに、本発明によれば、目的の蛋白質または生理活性ポリペプチドのアミノ酸配列データと、結合候補となる1つまたは複数の蛋白質または生理活性ポリペプチドのアミノ酸配列データとを取得し、目的の蛋白質または生理活性ポリペプチドと、結合候補となる蛋白質または生理活性ポリペプチドとが結合した複合体の立体構造情報を生成し、取得された目的のアミノ酸配列データと結合候補のアミノ酸配列データに含まれる各アミノ酸残基の間の空間上の距離を、生成した複合体の立体構造情報に従って決定し、目的のアミノ酸配列データと結合候補のアミノ酸配列データに含まれる各アミノ酸残基が持つ電荷を決定し、  
15 決定された各アミノ酸残基間の空間上の距離と、決定された各アミノ酸残基が持つ電荷とに従って、各アミノ酸残基のエネルギーを計算し、複合体について結合部位を変えて複合体の立体構造情報を生成し、各アミノ酸残基のエネルギーを計算し、当該エネルギーの総和が最小となる結合部位を求め、全ての結合候補についてエネルギー最小化処理を実行した結果、エネルギーの総和が最小  
20 となる結合部位をもつ結合候補を決定するので、蛋白質または生理活性ポリペプチドのアミノ酸配列から静電的に不安定になりそうなアミノ酸残基が結合部位になりやすいことを利用して、高速かつ高精度に最適な結合候補の蛋白質を

予測することができる結合部位予測装置、結合部位予測方法、プログラム、および、記録媒体を提供することができる。

(V) また、以下に、本発明にかかるタンパク質構造最適化装置、タンパク質構造最適化方法、プログラム、および、記録媒体の実施の形態を図面に基づいて詳細に説明する。なお、この実施の形態によりこの発明が限定されるものではない。

特に、以下の実施の形態においては、本発明を、富士通（会社名）の「MOPAC 2000 ver. 1.0」（製品名）に適用した例について説明するが、この場合に限られず、他のプログラムを用いても同様に適用することができる。

#### [本発明の概要]

以下、本発明の概要について説明し、その後、本発明の構成および処理等について詳細に説明する。図72は本発明の基本原理を示すフローチャートである。

本発明は、概略的に、以下の基本的特徴を有する。まず、本発明は、タンパク質の座標データを取得する（ステップSA4-1）。ここで、取得するタンパク質の座標データは、例えば、X線結晶解析により求めた座標データに既知のモデリングソフト（例えば、Accelrys Inc.（会社名）の「WebLab Viewer Pro 4.2」（製品名）や「Insight II」（製品名）（[www.accelrys.com](http://www.accelrys.com)）、Tripos, Inc.（会社名）の「SYBYL 6.7」（製品名）（[www.tripos.com](http://www.tripos.com)）、CambridgeSoft Corporation（会社名）の「Chem3D 7.0」（製品名）（[www.camsoft.com](http://www.camsoft.com)）など）を用いて水素を付加したものや、また、PDB（Protein Data Base）などの既知の蛋白構造データベースに登録された座標データなど、あらゆるタンパク質の座標データを用いてもよい。



そして、本発明は、タンパク質の座標データについて、特定のアミノ酸残基  $i$  から所定の距離（例えば、 $r$  Å）内に含まれる近傍アミノ酸残基群の座標を抽出する（ステップ S A 4-2）。すなわち、アミノ酸残基  $i$  に含まれる全原子から所定の距離内に存在する原子を含むアミノ酸残基の群が近傍アミノ酸残基群であり、この近傍アミノ酸残基群に含まれる全原子座標を抽出する。抽出した近傍アミノ酸残基群の中にシステイン（CYS）が含まれており、それが別のシステイン（CYS）とジスルフィド結合をしている場合には、その別の CYS も近傍アミノ酸残基群に加えてもよい。

ステップ S A 4-2 の操作で機械的に座標を切り取った場合、その切り口がラジカルとなり計算に不都合が生じる。これを解消するために、本発明は、近傍アミノ酸残基群の切り口の部分にキャップ用置換基（例えば水素原子（H）またはメチル基（CH<sub>3</sub>））を付加する（ステップ S A 4-3）。

そして、本発明は、キャップ用置換基が付加された近傍アミノ酸残基群の全体の電荷を計算する（ステップ S A 4-4）。電荷計算は、既知のあらゆる電荷計算手法を用いてもよいが、例えば、塩基性アミノ酸残基数から酸性アミノ酸残基数を減算することにより高速に計算することもできる。

そして、本発明は、キャップ用置換基が付加された近傍アミノ酸残基群について、電荷を用いて、特定のアミノ酸残基  $i$  の原子座標について既知の分子軌道計算プログラム（例えば、「MOPAC 2000 ver. 1.0」（製品名）などの半経験的分子軌道計算プログラム）などを用いて構造最適化を実行する（ステップ S A 4-5）。

そして、本発明は、最適化された原子座標を、初期のタンパク質の座標データ上の対応する原子座標と置換する（ステップ S A 4-6）。

そして、本発明は、全てのアミノ酸残基  $i$  についてステップ S A 4-2 からステップ S A 4-6 を適用し（1 番目のアミノ酸残基から最後のアミノ酸残基まで  $i$  をインクリメントしてループ処理を行う。）、全てのアミノ酸残基について最適化を実行する（ステップ S A 4-7）。

そして、本発明は、ステップS A 4－7において得られた構造データを初期構造として、ステップS A 4－1からステップS A 4－7を複数回（n回）実行することにより、更に構造最適化の精度を上げることができる（ステップS A 4－8）。

5     [システム構成]

まず、本システムの構成について説明する。図73は、本発明が適用される本システムの構成の一例を示すブロック図であり、該構成のうち本発明に係る部分のみを概念的に示している。本システムは、概略的に、タンパク質構造最適化装置4100と、タンパク質構造情報等に関する外部データベースや  
10     ホモロジー検索等の外部プログラム等を提供する外部システム4200とを、ネットワーク4300を介して通信可能に接続して構成されている。

図73においてネットワーク4300は、タンパク質構造最適化装置4100と外部システム4200とを相互に接続する機能を有し、例えば、インターネット等である。

15     図73において外部システム4200は、ネットワーク4300を介して、タンパク質構造最適化装置4100と相互に接続され、利用者に対してタンパク質構造情報等に関する外部データベースやホモロジー検索やモチーフ検索等の外部プログラムを実行するウェブサイトを提供する機能を有する。

ここで、外部システム4200は、WEBサーバやASPサーバ等として構成してもよく、そのハードウェア構成は、一般に市販されるワークステーション、パーソナルコンピュータ等の情報処理装置およびその付属装置により構成してもよい。また、外部システム4200の各機能は、外部システム4200  
20     のハードウェア構成中のCPU、ディスク装置、メモリ装置、入力装置、出力装置、通信制御装置等およびそれらを制御するプログラム等により実現される。

25     図73においてタンパク質構造最適化装置4100は、概略的に、タンパク質構造最適化装置4100の全体を統括的に制御するCPU等の制御部4102、通信回線等に接続されるルータ等の通信装置（図示せず）に接続される通

信制御インターフェース部 4 1 0 4、入力装置 4 1 1 2 や出力装置 4 1 1 4 に  
接続される入出力制御インターフェース部 4 1 0 8、および、各種のデータベ  
ースやテーブルなどを格納する記憶部 4 1 0 6 を備えて構成されており、これ  
ら各部は任意の通信路を介して通信可能に接続されている。さらに、このタン  
5 パク質構造最適化装置 4 1 0 0 は、ルータ等の通信装置および専用線等の有線  
または無線の通信回線を介して、ネットワーク 4 3 0 0 に通信可能に接続され  
ている。

記憶部 4 1 0 6 に格納される各種のデータベースやテーブルなど（タンパク  
質構造情報データベース 4 1 0 6 a および処理結果ファイル 4 1 0 6 b）は、  
10 固定ディスク装置等のストレージ手段であり、各種処理に用いる各種のプログ  
ラムやテーブルやファイルやデータベースやウェブページ用ファイル等を格納  
する。

これら記憶部 4 1 0 6 の各構成要素のうち、タンパク質構造情報データベ  
ース 4 1 0 6 a は、タンパク質の立体構造の座標データ等を格納した座標データ  
15 格納手段である。タンパク質構造情報データベース 4 1 0 6 a は、インターネ  
ットを経由してアクセスする P D B などの外部のデータベースであってもよく、  
また、これらのデータベースをコピーしたり、オリジナルの情報を格納したり、  
さらに独自のアノテーション情報等を付加したりして作成したインハウスデー  
タベースであってもよい。

20 また、処理結果ファイル 4 1 0 6 b は、タンパク質構造最適化装置 4 1 0 0  
の制御部 4 1 0 2 の各処理による処理結果に関する情報等を格納する処理結果  
格納手段である。

また、図 7 3 において、通信制御インターフェース部 4 1 0 4 は、タンパク  
質構造最適化装置 4 1 0 0 とネットワーク 4 3 0 0（またはルータ等の通信装  
25 置）との間における通信制御を行う。すなわち、通信制御インターフェース部  
4 1 0 4 は、他の端末と通信回線を介してデータを通信する機能を有する。

また、図 7 3 において、入出力制御インターフェース部 4 1 0 8 は、入力装置 4 1 1 2 や出力装置 4 1 1 4 の制御を行う。ここで、出力装置 4 1 1 4 としては、モニタ（家庭用テレビを含む）の他、スピーカを用いることができる（なお、以下においては出力装置 4 1 1 4 をモニタとして記載する場合がある）。また、入力装置 4 1 1 2 としては、キーボード、マウス、および、マイク等を用いることができる。また、モニタも、マウスと協働してポインティングデバイス機能を実現する。

また、図 7 3 において、制御部 4 1 0 2 は、OS (Operating System) 等の制御プログラム、各種の処理手順等を規定したプログラム、および所要データを格納するための内部メモリを有し、これらのプログラム等により、種々の処理を実行するための情報処理を行う。制御部 4 1 0 2 は、機能概念的に、座標データ取得部 4 1 0 2 a、近傍アミノ酸残基群抽出部 4 1 0 2 b、キャップ付加部 4 1 0 2 c、電荷計算部 4 1 0 2 d、構造最適化部 4 1 0 2 e、および、原子座標置換部 4 1 0 2 f を備えて構成されている。

このうち、座標データ取得部 4 1 0 2 a は、タンパク質の座標データを取得する座標データ取得手段である。また、近傍アミノ酸残基群抽出部 4 1 0 2 b は、タンパク質の座標データについて、特定のアミノ酸残基から所定の距離内に含まれる近傍アミノ酸残基群の座標を抽出する近傍アミノ酸残基群抽出手段である。また、キャップ付加部 4 1 0 2 c は、近傍アミノ酸残基群の切り口の部分にキャップ用置換基を付加するキャップ付加手段である。また、電荷計算部 4 1 0 2 d は、キャップ付加手段により上記キャップ用置換基が付加された上記近傍アミノ酸残基群の全体の電荷を計算する電荷計算手段である。また、構造最適化部 4 1 0 2 e は、キャップ付加手段により上記キャップ用置換基が付加された上記近傍アミノ酸残基群について、上記電荷計算手段により計算された上記電荷を用いて上記特定のアミノ酸残基の原子座標について構造最適化を実行する構造最適化手段である。また、原子座標置換部 4 1 0 2 f は、構造最適化手段にて最適化された上記原子座標を、上記タンパク質の座標データ上

の対応する原子座標と置換する原子座標置換手段である。なお、これら各部によって行なわれる処理の詳細については、後述する。

〔システムの処理〕

次に、このように構成された本実施の形態における本システムの処理の一例について、以下に図 7 4～図 9 0 を参照して詳細に説明する。

〔メイン処理〕

まず、メイン処理の詳細について図 7 4 を参照して説明する。図 7 4 は、本実施形態における本システムのメイン処理の一例を示すフローチャートである。

タンパク質構造最適化装置 4 1 0 0 は、座標データ取得部 4 1 0 2 a の処理により、タンパク質構造情報データベース 4 1 0 6 a や外部システム 4 2 0 0 の外部データベースから所望のタンパク質の座標データを取得する（ステップ S B 4 - 1）。ここで、取得するタンパク質の座標データは、例えば、X線結晶解析により求めた座標データに既知のモデリングソフト（例えば、Accelerlys Inc.（会社名）の「WebLab Viewer Pro 4.2」（製品名）や「Insight II」（製品名）（www.accellys.com）、Tripos, Inc.（会社名）の「SYBYL 6.7」（製品名）（www.tripos.com）、CambridgeSoft Corporation（会社名）の「Chem3D 7.0」（製品名）（www.camsoft.com）など）を用いて水素を付加したものや、また、PDB（Protein Data Bank）などの既知の蛋白構造データベースに登録された座標データなど、あらゆるタンパク質の座標データを用いてもよい。

ここで、図 7 5 は、タンパク質の座標データの一例を示す図である。図 7 5 に示す例では、PDB形式の座標データを用いており、また、X線結晶解析で求めた構造情報に市販のプログラムにより水素を付加している。

再び図 7 4 に戻り、タンパク質構造最適化装置 4 1 0 0 は、制御部 4 1 0 2 の処理により、処理回数を表すカウンタ n（初期値は 0）に 1 を加算する

(ステップS B 4-2)。

また、タンパク質構造最適化装置4100は、制御部4102の処理により、アミノ酸残基番号を表すカウンター*i* (初期値は0) に1を加算する (ステップS B 4-3)。

- 5       そして、タンパク質構造最適化装置4100は、近傍アミノ酸残基群抽出部4102bの処理により、処理対象のタンパク質の座標データについて、特定のアミノ酸残基*i* から所定の距離 (例えば、*r* Å) 内に含まれる近傍アミノ酸残基群の座標を抽出する (ステップS B 4-4)。すなわち、アミノ酸残基*i* に含まれる全ての原子*j* から所定の距離内に存在する原子1を含むアミノ酸残基
- 10       基*k* (*k* ≠ *i*) の群が近傍アミノ酸残基群であり、この近傍アミノ酸残基群に含まれる全ての原子*m*の座標を抽出する。

- また、抽出した近傍アミノ酸残基群の中にシステイン (CYS) が含まれており、それが別のシステイン (CYS) とジスルフィド結合をしている場合には、その別のシステイン (CYS) も近傍アミノ酸残基群に加えてもよい。す
- 15       なわち、近傍アミノ酸残基群抽出部4102bは、抽出した近傍アミノ酸残基群の中にシステイン (CYS) が含まれている場合には、当該システイン (CYS) とジスルフィド結合をしている近傍アミノ酸残基群には含まれない別のシステイン (CYS) が存在しているか判定し、当該別のシステイン (CYS) が存在する場合には当該別のシステイン (CYS) も近傍アミノ酸残基群
- 20       に加える。

- ステップS B 4-4の操作で機械的に座標を切り取った場合、その切り口がラジカルとなり計算に不都合が生じる。これを解消するために、タンパク質構造最適化装置4100は、キャップ付加部4102cの処理により、近傍アミノ酸残基群の切り口の部分にキャップ用置換基 (例えば水素原子 (H) または
- 25       メチル基 (CH<sub>3</sub>) ) を付加する (ステップS B 4-5)。キャップ用置換基として水素、メチル基のどちらを用いるかは、ユーザが目的に応じて判断する。

ここで、キャップ付加部 4 1 0 2 c のよるキャップ付加処理の詳細について図 7 6 から図 8 3 を参照して説明する。

図 7 6 は、本実施形態における切り口に水素原子を付加する場合のキャップ付加処理の一例を示すフローチャートである。また、図 7 7 は、もとの座標と  
5 キャップ置換基を付加した後の座標の概念を示す図である。図 7 6 は、図 7 7 に示すもとの座標（左側）に対してアミノ基側へキャップ付加（右側）をする際の処理の一例を示す。近傍アミノ酸残基群の任意の残基を  $j$  とする。

キャップ付加部 4 1 0 2 c は、アミノ酸残基  $j$  が N 端アミノ酸である場合（ステップ S C 4 - 1）、アミノ酸残基  $j$  のアミノ側は切り口になっていない  
10 ので、キャップ付加は不要とする（ステップ S C 4 - 2）。

そして、キャップ付加部 4 1 0 2 c は、アミノ酸残基  $j$  が N 端アミノ酸でない場合には（ステップ S C 4 - 1）、隣のアミノ酸残基  $j - 1$  も抽出されたアミノ酸残基群に含まれる場合（ステップ S C 4 - 3）、残基  $j$  のアミノ側は切り口になっていないので、キャップ付加は不要とする（ステップ S C 4 - 4）。

一方、キャップ付加部 4 1 0 2 c は、隣のアミノ酸残基  $j - 1$  も抽出されたアミノ酸残基群に含まれない場合には（ステップ S C 4 - 3）、アミノ酸残基  
15  $j - 1$  の主鎖カルボニル炭素を  $C_{j-1}$  とする（ステップ S C 4 - 5）。

そして、キャップ付加部 4 1 0 2 c は、アミノ酸残基  $j$  の主鎖アミノ基窒素を  $N_j$  とする（ステップ S C 4 - 6）。

20 そして、キャップ付加部 4 1 0 2 c は、付加するキャップ水素原子  $H_{CAPN}$  の位置を以下の数式（1）に従って決める（ステップ S C 4 - 7）。

$$\overrightarrow{N_j H_{CAPN}} = \frac{\overrightarrow{N_j C_{j-1}}}{\left| \overrightarrow{N_j C_{j-1}} \right|} \times R_{NH} \quad (R_{NH} = 1.01 \text{Å}) \quad \text{数式 (1)}$$

25 次に図 7 8 は、本実施形態における切り口に水素原子を付加する場合のキャ

アップ付加処理の一例を示すフローチャートである。また、図 7 9 は、もとの座標とキャップ置換基を付加した後の座標の概念を示す図である。図 7 8 は、図 7 9 に示すもとの座標（左側）に対してカルボキシル基側へキャップ付加（右側）をする際の処理の一例を示す。近傍アミノ酸残基群の任意の残基を  $j$  とする。

キャップ付加部 4 1 0 2 c は、アミノ酸残基  $j$  が C 端アミノ酸である場合（ステップ S D 4-1）、アミノ酸残基  $j$  のアミノ側は切り口になっていないので、キャップ付加は不要とする（ステップ S D 4-2）。

そして、キャップ付加部 4 1 0 2 c は、アミノ酸残基  $j$  が C 端アミノ酸でない場合には（ステップ S D 4-1）、隣のアミノ酸残基  $j+1$  も抽出されたアミノ酸残基群に含まれる場合（ステップ S D 4-3）、残基  $j$  のアミノ側は切り口になっていないので、キャップ付加は不要とする（ステップ S D 4-4）。

一方、キャップ付加部 4 1 0 2 c は、隣のアミノ酸残基  $j+1$  も抽出されたアミノ酸残基群に含まれない場合には（ステップ S D 4-3）、アミノ酸残基  $j+1$  の主鎖アミノ基窒素を  $N_{j+1}$  とする（ステップ S D 4-5）。

そして、キャップ付加部 4 1 0 2 c は、アミノ酸残基  $j$  の主鎖カルボニル炭素を  $C_j$  とする（ステップ S D 4-6）。

そして、キャップ付加部 4 1 0 2 c は、付加するキャップ水素原子  $H_{CAPC}$  の位置を以下の数式（2）に従って決める（ステップ S D 4-7）。

$$\overrightarrow{C_j H_{CAPC}} = \frac{\overrightarrow{C_j N_{j+1}}}{\left| \overrightarrow{C_j N_{j+1}} \right|} \times R_{Csp2H} \quad (R_{Csp2H} = 1.08 \text{Å}) \quad \text{数式 (2)}$$

次に図 8 0 は、本実施形態における切り口にメチル基を付加する場合のキャップ付加処理の一例を示すフローチャートである。また、図 8 1 は、もとの座標とキャップ置換基を付加した後の座標の概念を示す図である。図 8 0 は、図



8 1に示すもとの座標（左側）に対してアミノ基側へキャップ付加（右側）をする際の処理の一例を示す。近傍アミノ酸残基群の任意の残基を  $j$  とする。

キャップ付加部 4 1 0 2 c は、アミノ酸残基  $j$  が N 端アミノ酸である場合（ステップ S E 4 - 1）、アミノ酸残基  $j$  のアミノ側は切り口になっていない  
5 ので、キャップ付加は不要とする（ステップ S E 4 - 2）。

そして、キャップ付加部 4 1 0 2 c は、アミノ酸残基  $j$  が N 端アミノ酸でない場合には（ステップ S E 4 - 1）、隣のアミノ酸残基  $j - 1$  も抽出されたアミノ酸残基群に含まれる場合（ステップ S E 4 - 3）、残基  $j$  のアミノ側は切り口になっていないので、キャップ付加は不要とする（ステップ S E 4 - 4）。  
10 一方、キャップ付加部 4 1 0 2 c は、隣のアミノ酸残基  $j - 1$  も抽出されたアミノ酸残基群に含まれない場合には（ステップ S E 4 - 3）、アミノ酸残基  $j - 1$  の主鎖カルボニル炭素を  $C_{j-1}$  とする（ステップ S E 4 - 5）。

そして、キャップ付加部 4 1 0 2 c は、アミノ酸残基  $j$  の主鎖アミノ基窒素を  $N_j$  とする（ステップ S E 4 - 6）。

15 そして、キャップ付加部 4 1 0 2 c は、アミノ酸残基  $j$  の主鎖  $\alpha$  炭素を  $C A_j$  とする（ステップ S E 4 - 7）。

そして、キャップ付加部 4 1 0 2 c は、付加するキャップメチル基炭素  $C_{CAPN}$  の位置を以下の数式（3）に従って決める（ステップ S E 4 - 8）。

$$20 \quad \overrightarrow{N_j C_{CAPN}} = \frac{\overrightarrow{N_j C_{j-1}}}{\left| \overrightarrow{N_j C_{j-1}} \right|} \times R_{NCsp3} \quad (R_{NCsp3} = 1.47 \text{ \AA}) \quad \text{数式 (3)}$$

そして、キャップ付加部 4 1 0 2 c は、付加する 3 つのキャップメチル基水素  $H_{CAPNk}$  ( $k = 1, 2, 3$ ) の位置を以下の条件（数式（4））で決める（ステップ S E 4 - 9）。

$$\begin{aligned}
 \text{結合長} \quad & \left| \overrightarrow{H_{CAPNk} C_{CAPN}} \right| = R_{Csp3H} \quad (R_{Csp3H} = 1.09 \text{ \AA}) \\
 \text{結合角} \quad & \angle H_{CAPNk} C_{CAPN} N_j = A_{Csp3} \quad (A_{Csp3} = 109.5^\circ) \\
 \text{二面角} \quad & \angle H_{CAPNk} C_{CAPN} N_j C A_j = D_k \quad (D_1 = 180.0^\circ, D_2 = 60.0^\circ, D_3 = -60.0^\circ)
 \end{aligned}$$

数式 (4)

次に図 8 2 は、本実施形態における切り口にメチル基を付加する場合のキャップ付加処理の一例を示すフローチャートである。また、図 8 3 は、もとの座標とキャップ置換基を付加した後の座標の概念を示す図である。図 8 2 は、図 8 3 に示すもとの座標（左側）に対してカルボキシル基側へキャップ付加（右側）をする際の処理の一例を示す。近傍アミノ酸残基群の任意の残基を  $j$  とする。

5      キャップ付加部 4 1 0 2 c は、アミノ酸残基  $j$  が C 端アミノ酸である場合（ステップ S F 4-1）、アミノ酸残基  $j$  のアミノ側は切り口になっていないので、キャップ付加は不要とする（ステップ S F 4-2）。

そして、キャップ付加部 4 1 0 2 c は、アミノ酸残基  $j$  が C 端アミノ酸でない場合には（ステップ S F 4-1）、隣のアミノ酸残基  $j+1$  も抽出されたアミノ酸残基群に含まれる場合（ステップ S F 4-3）、残基  $j$  のアミノ側は切り口になっていないので、キャップ付加は不要とする（ステップ S F 4-4）。

一方、キャップ付加部 4 1 0 2 c は、隣のアミノ酸残基  $j+1$  も抽出されたアミノ酸残基群に含まれない場合には（ステップ S F 4-3）、アミノ酸残基  $j+1$  の主鎖アミノ基窒素を  $N_{j+1}$  とする（ステップ S F 4-5）。

そして、キャップ付加部 4 1 0 2 c は、アミノ酸残基  $j$  の主鎖カルボニル炭素を  $C_j$  とする（ステップ S F 4-6）。

そして、キャップ付加部 4 1 0 2 c は、アミノ酸残基  $j$  の主鎖  $\alpha$  炭素を  $C A_j$  とする（ステップ S F 4-7）。

そして、キャップ付加部 4 1 0 2 c は、付加するキャップメチル基炭素  $C$

$C_{APC}$  の位置を以下の数式 (5) に従って決める (ステップ S F 4-8)。

$$\overrightarrow{C_j C_{APC}} = \frac{\overrightarrow{C_j N_{j+1}}}{|\overrightarrow{C_j N_{j+1}}|} \times R_{Csp2Csp3} \quad (R_{Csp2Csp3} = 1.52) \quad \text{数式 (5)}$$

- 5      そして、キャップ付加部 4 1 0 2 c は、付加する 3 つのキャップメチル基水素  $H_{CAPCk}$  ( $k = 1, 2, 3$ ) の位置を以下の条件 (数式 (6)) で決める (ステップ S F 4-9)。

$$\begin{aligned} \text{結合長} \quad |\overrightarrow{H_{CAPCk} C_{APC}}| &= R_{Csp3H} \quad (R_{Csp3H} = 1.09 \text{ \AA}) \\ \text{結合角} \quad \angle H_{CAPCk} C_{APC} C_j &= A_{Csp3} \quad (A_{Csp3} = 109.5^\circ) \\ \text{二面角} \quad \angle H_{CAPCk} C_{APC} C_j C A_j &= D_k \quad (D_1 = 180.0^\circ, D_2 = 60.0^\circ, D_3 = -60.0^\circ) \\ &\text{数式 (6)} \end{aligned}$$

10

ここで、数式 (1) ~ 数式 (6) において、R、A、D は、それぞれ、標準結合長、標準結合角、標準二面角であり、本条件で記した数値はその一例である (平野恒夫・田辺和俊編「分子軌道法 MOPAC ガイドブック (3 訂版)」海文堂出版、1999 参照)。

- 15      これにて、キャップ付加処理が終了する。

再び図 7 4 に戻り、タンパク質構造最適化装置 4 1 0 0 は、全ての近傍アミノ酸残基群の切口にキャップを付加すると、ステップ S B 4-4 で抽出したアミノ酸残基群全体の電荷計算を行う。すなわち、MOPAC 2000 に限らず、一般に分子軌道計算を行う際には対象となる系全体の電荷を入力データとして与えるため、タンパク質構造最適化装置 4 1 0 0 は、電荷計算部 4 1 0 2 d の処理により、キャップ用置換基が付加された近傍アミノ酸残基群の全体の電荷を計算する (ステップ S B 4-6)。

20

電荷計算は、既知のあらゆる電荷計算手法を用いてもよいが、例えば、以下の数式（7）を用いて、塩基性アミノ酸残基数から酸性アミノ酸残基数を減算することにより高速に計算することもできる。

$$\begin{aligned} 5 \quad & (\text{全体電荷}) = (\text{塩基性アミノ酸残基数}) - (\text{酸性アミノ酸残基数}) \\ & \dots \text{数式 (7)} \end{aligned}$$

ここで、塩基性アミノ酸残基は、ARG、LYSなどであり、酸性アミノ酸残基は、ASP、GLUなどである。アミノ酸種別は、図84に示すように、  
10 入力データとして与えるPDB形式データの三文字表記（18－20カラムの文字）で判別する（「PDB File Format Contents Guide Version 2.2」（20 December 1996）参照）。また、中性アミノ酸残基（例えば、ARG、LYS、ASP、GLUなど）や、プロトン化したHIS（電荷+1）の表記法は、分子動力学計算プログラム「Amber 7」（University of California, 2002.）の様式に従いARN, LYN, ASH, GLH, HIPとして、入力PDBデータで記述し判別する。また、非天然のアミノ酸残基や、ユーザ定義のアミノ酸、リガンド分子の電荷も個別に設定できるようにしておく。例えば、リン酸化したTHRをTPOと定義し、このアミノ酸に対しては-2の電荷を与えるようにプログラムで設定しておく。  
20

そして、タンパク質構造最適化装置4100は、構造最適化部4102eの処理により、MOPAC2000の入力ファイルを作成するために、アミノ酸残基iを構成する各原子に対して最適化処理対象の原子であることを表す「最適化フラグ」を設定する（ステップSB4-7）。なお、MOPAC2000  
25 に限らず、一般に計算化学的手法（分子軌道法、分子力学法など）で構造最適化を行う際には、最適な位置に動かす原子と、座標を固定して位置を動かさない原子を設定し、部分構造最適化を行うこともできる。ここでは、最適な位置

に動かす原子を入力データとして判別できるように設定する事を、MOPAC 2000の慣例に倣って、「最適化フラグを設定する」と呼ぶことにする（「MOPAC 2000 Manual」Fujitsu Limited, Tokyo, 2000 参照）。

- 5        具体的には、構造最適化部 4102e は、水素の構造最適化を行う場合、アミノ酸残基  $i$  の水素原子に最適化フラグを設定する。図 85 は、アミノ酸残基  $i$  の水素原子に最適化フラグを設定する場合の一例を示す図である。図 85 は、PDBコードが「1CBI」のタンパク質に対して水素付加を行った入力 PDBデータに対して、特定アミノ酸残基が 50 番目のアミノ酸残基 ( $i = 50$ ) であり、距離が 3.0 オングストローム ( $r = 3.0 \text{ \AA}$ ) のときに抽出される近傍アミノ酸残基群を示している。また、上述の方法により近傍アミノ酸残基群の切り口にキャップ置換基（水素原子）の付加も行っている。また、上述のステップ SB4-6 では、ここに示した全原子を考慮して電荷計算を行っている。図 85 において、太線と球で表示されている部分が計算の中心残基である PHE50 ( $i = 50$  のアミノ酸残基であるフェニルアラニン) である。この PHE50 の中で最適化フラグを設定する水素原子を球で示している。
- 10
- 15

- また、構造最適化部 4102e は、側鎖の構造最適化を行う場合、アミノ酸残基  $i$  の水素と側鎖原子に最適化フラグを設定する。図 86 は、アミノ酸残基  $i$  の水素と側鎖原子に最適化フラグを設定する場合の一例を示す図である。図 86 は、PDBコードが「1CBI」のタンパク質に対して水素付加を行った入力 PDBデータに対して、特定アミノ酸残基が 50 番目のアミノ酸残基 ( $i = 50$ ) であり、距離が 3.0 オングストローム ( $r = 3.0 \text{ \AA}$ ) のときに抽出される近傍アミノ酸残基群を示している。また、上述の方法により近傍アミノ酸残基群の切り口にキャップ置換基（水素原子）の付加も行っている。また、上述のステップ SB4-6 では、ここに示した全原子を考慮して電荷計算を行っている。図 86 において、太線と球で表示されている部分が計算の中心残基である PHE50 ( $i = 50$  のアミノ酸残基であるフェニルアラニン) である。
- 20
- 25

このPHE 50の中で最適化フラグを設定する水素原子と側鎖原子を球で示している。

また、構造最適化部4102eは、全原子の構造最適化を行う場合、アミノ酸残基iの全原子に最適化フラグを設定する。ただし、MOPAC2000を含め現状の分子軌道理論では、主鎖構造の2次構造を再現することが困難なため、主鎖原子の最適化は一般には行わないことが多いが2次構造を再現できる精度の高い理論が構築されれば、全構造最適化も有効となる。

再び図74に戻り、タンパク質構造最適化装置4100は、構造最適化部4102eの処理により、MOPAC2000の入力ファイルを作成する（ステップSB4-8）。図87は、MOPAC2000の入力ファイルの一例を示す図である。図87に示すように、電荷、近傍アミノ酸残基群の座標データ、最適化フラグなどを含む入力ファイルを作成する。

そして、タンパク質構造最適化装置4100は、構造最適化部4102eの処理により、キャップ用置換基が付加された近傍アミノ酸残基群について、電荷を用いて、特定のアミノ酸残基の原子座標についてMOPAC2000を用いて構造最適化を実行する（ステップSB4-9）。ここで、図88は、MOPAC2000による構造最適化処理の結果を示す出力ファイルの一例を示す図である。図88に示すように、構造最適化後の座標データが出力される。なお、図88において、「\*」の付いた座標が最適化された部分を示している。

そして、タンパク質構造最適化装置4100は、原子座標置換部4102fの処理により、最適化された原子座標を、初期のタンパク質の座標データ上の対応する原子座標と置換する（ステップSB4-10）。すなわち、原子座標置換部4102fは、MOPAC2000の処理結果（出力ファイル）の

「\*」の付いた座標が最適化部分であるため、この部分を抜き出しステップSB4-1で用意した座標データの対応する座標部分と置き換える。

そして、タンパク質構造最適化装置4100は、全てのアミノ酸残基iについてステップSB4-3からステップSB4-10を適用し（1番目のアミノ

酸残基から最後のアミノ酸残基まで  $i$  をインクリメントしてループ処理を行う。) 、全てのアミノ酸残基について最適化を実行する (ステップ S B 4-11) 。

そして、タンパク質構造最適化装置 4100 は、ステップ S B 4-10 において得られた構造データを初期構造として、ステップ S B 4-2 からステップ S B 4-7 を複数回 ( $n$  回) 実行することにより、更に構造最適化の精度を上げることができる (ステップ S B 4-12) 。すなわち、ステップ S B 4-4 からステップ S B 4-10 の処理を N 残基から C 端残基まで行うことで、すべてのアミノ酸残基について部分構造が最適化された P D B 形式の座標データが得られる。この座標データを入力として、座標は固定して (すべての原子に対して最適化フラグを設定せずに) M O P A C でエネルギー計算を行う。また、上述のステップ S B 4-4 からステップ S B 4-10 の操作を含むループ処理は、例えばスクリプトプログラムを用いて実行してもよい。

これにて、メイン処理が終了する。

#### 15 [本発明による計算例]

次に、本発明による計算例の詳細について図 89 および図 90 などを参照して説明する。本計算例では、サンプル分子として「J a p a n e s e P e a r S 3-R i b o n u c l e a s e」 (P D B I D : 1 I Q Q A) を用いて、特定アミノ酸残基を 200 番目のアミノ酸残基 (3262 原子 C1047 H1619 N285 O300 S11) をした場合である。また、本計算例で使用した計算機の機種名は、COMPAQ (会社名) の「A l p h a S e r v e r E S 4 0 (C P U A l p h a 2 1 2 6 4 8 3 3 M H z)」 (製品名) である。図 89 は、従来の最適化手法 (M O Z Y M E 法+B F G S 法) と本発明の手法により水素構造を最適化した場合の計算結果を示す図である。また、図 90 は、従来の最適化手法 (M O Z Y M E 法+B F G S 法) と本発明の手法により側鎖構造を最適化した場合の計算結果を示す図である。図 89 および図 90 において、縦軸は、生成熱 H e a t o f F o r m a t i o

$n$  (kcal mol<sup>-1</sup>) であり、横軸はCPU時間 (秒) を示している。また、初期構造のHeat of Formationの値は、-1044.53571 kcal・mol<sup>-1</sup>である。

本計算例について計算時間とエネルギー (生成熱) の関係をみると、本発明  
5 の手法では、計算時間に対してエネルギーの収束が速く、全体のループを3～5回繰り返すこと ( $n=3\sim5$ ) で、エネルギーが収束することがわかる。また、計算精度よりも計算時間を優先する場合には、 $r$ の値を小さめにとり、その逆に計算精度を求める場合には、 $r$ の値を大きめにとることもできる。

また、本計算例について必要となる最大メモリ容量をみると、従来の手法では、水素構造最適化を行う場合は、506MB (メガバイト) であり、側鎖構造最適化を行う場合は、667MBであった。一方、本発明の手法では、水素構造最適化を行う場合は、301MBであり、側鎖構造最適化を行う場合は、301MBであった。このように、本発明の手法ではメモリの少量化も図ることができた。

#### 15 [他の実施の形態]

さて、これまで本発明の実施の形態について説明したが、本発明は、上述した実施の形態以外にも、上記特許請求の範囲に記載した技術的思想の範囲内において種々の異なる実施の形態にて実施されてよいものである。

例えば、タンパク質構造最適化装置4100がスタンドアローンの形態で処理を行う場合を一例に説明したが、タンパク質構造最適化装置4100とは別筐体で構成されるクライアント端末からの要求に応じて処理を行い、その処理結果を当該クライアント端末に返却するように構成してもよい。

また、上述した実施の形態では、半経験的分子軌道プログラムであるMOPAC2000を用いる場合を一例として説明したが、他の既知の計算手法やプログラムを用いてもよい。例えば、分子軌道計算プログラムである「Gaussian 98 Rev. A. 11. 3」 (製品名) (Gaussian, Inc. (会社名), Pittsburgh PA, 2002)、「Game



ss June 20 2002 R2」(製品名)(Iowa State University, 2002)などのプログラムに置き換えれば、アブイニシオ(ab initio)分子軌道法による構造最適化が可能になる。また、「Amber 7」(製品名)(University of California, 2002)や、「Tinker 3.7」(製品名)(Washington University School of Medicine, 2001)などに置き換えれば、分子力学計算の高速化も可能である。これらのプログラムの入出力データは、MOPAC 2000の入力ファイルと座標パラメータの並び方などが異なるだけであるため、「Babel version 1.6」(製品名)(Pat Walters and Matt Stahl, 1996)等のプログラムを用いて容易にMOPAC 2000の入出力データと変換することができる。MOPAC 2000は半経験的分子軌道プログラムと呼ばれ半定量的な結果が得られる。一方、GaussianやGamessなどは、ab initio分子軌道計算プログラムと呼ばれ半経験的な方法よりも定量的な結果が得られるが、計算時間は半経験的な方法より格段にかかるのが一般的である。

また、実施形態において説明した各処理のうち、自動的に行なわれるものとして説明した処理の全部または一部を手動的に行うこともでき、あるいは、手動的に行なわれるものとして説明した処理の全部または一部を公知の方法で自動的に行うこともできる。

この他、上記文書中や図面中で示した処理手順、制御手順、具体的名称、各種の登録データや検索条件等のパラメータを含む情報、画面例、データベース構成については、特記する場合を除いて任意に変更することができる。

また、タンパク質構造最適化装置4100に関して、図示の各構成要素は機能概念的なものであり、必ずしも物理的に図示の如く構成されていることを要しない。

例えば、タンパク質構造最適化装置4100の各部または各装置が備える処

理機能、特に制御部 4 1 0 2 にて行なわれる各処理機能については、その全部または任意の一部を、CPU (Central Processing Unit) および当該CPUにて解釈実行されるプログラムにて実現することができ、あるいは、ワイヤードロジックによるハードウェアとして実現することも可能である。なお、プログラムは、後述する記録媒体に記録されており、必要に応じてタンパク質構造最適化装置 4 1 0 0 に機械的に読み取られる。

すなわち、ROMまたはHDなどの記憶部 4 1 0 6 などには、OS (Operating System) と協働してCPUに命令を与え、各種処理を行うためのコンピュータプログラムが記録されている。このコンピュータプログラムは、RAM等にロードされることによって実行され、CPUと協働して制御部 4 1 0 2 を構成する。また、このコンピュータプログラムは、タンパク質構造最適化装置 4 1 0 0 に対して任意のネットワーク 4 3 0 0 を介して接続されたアプリケーションプログラムサーバに記録されてもよく、必要に応じてその全部または一部をダウンロードすることも可能である。

また、本発明にかかるプログラムを、コンピュータ読み取り可能な記録媒体に格納することもできる。ここで、この「記録媒体」とは、フレキシブルディスク、光磁気ディスク、ROM、EPROM、EEPROM、CD-ROM、MO、DVD等の任意の「可搬用の物理媒体」や、各種コンピュータシステムに内蔵されるROM、RAM、HD等の任意の「固定用の物理媒体」、あるいは、LAN、WAN、インターネットに代表されるネットワークを介してプログラムを送信する場合の通信回線や搬送波のように、短期にプログラムを保持する「通信媒体」を含むものとする。

また、「プログラム」とは、任意の言語や記述方法にて記述されたデータ処理方法であり、ソースコードやバイナリコード等の形式を問わない。なお、「プログラム」は必ずしも単一的に構成されるものに限られず、複数のモジュールやライブラリとして分散構成されるものや、OS (Operating System) に代表される別個のプログラムと協働してその機能を達成する

ものをも含む。なお、実施の形態に示した各装置において記録媒体を読み取るための具体的な構成、読み取り手順、あるいは、読み取り後のインストール手順等については、周知の構成や手順を用いることができる。

また、タンパク質構造最適化装置 4 1 0 0 は、さらなる構成要素として、マウス等の各種ポインティングデバイスやキーボードやイメージスキャナやデジタルタイザ等から成る入力装置（図示せず）、入力データのモニタに用いる表示装置（図示せず）、システムクロックを発生させるクロック発生部（図示せず）、および、各種処理結果その他のデータを出力するプリンタ等の出力装置（図示せず）を備えてもよく、また、入力装置、表示装置および出力装置は、それぞれ入出力インターフェースを介して制御部 4 1 0 2 に接続されてもよい。

記憶部 4 1 0 6 に格納される各種のデータベース等（タンパク質構造情報データベース 4 1 0 6 a および処理結果ファイル 4 1 0 6 b）は、RAM、ROM等のメモリ装置、ハードディスク等の固定ディスク装置、フレキシブルディスク、光ディスク等のストレージ手段であり、各種処理やウェブサイト提供に用いる各種のプログラムやテーブルやファイルやデータベースやウェブページ用ファイル等を格納する。

また、タンパク質構造最適化装置 4 1 0 0 は、既知のパーソナルコンピュータ、ワークステーション等の情報処理端末等の情報処理装置にプリンタやモニタやイメージスキャナ等の周辺装置を接続し、該情報処理装置に本発明の方法を実現させるソフトウェア（プログラム、データ等を含む）を実装することにより実現してもよい。

さらに、タンパク質構造最適化装置 4 1 0 0 の分散・統合の具体的形態は図示のものに限られず、その全部または一部を、各種の負荷等に応じた任意の単位で、機能的または物理的に分散・統合して構成することができる。例えば、各データベースを独立したデータベース装置として独立に構成してもよく、また、処理の一部を CGI（Common Gateway Interface）を用いて実現してもよい。

また、ネットワーク 4300 は、タンパク質構造最適化装置 4100 と外部システム 4200 とを相互に接続する機能を有し、例えば、インターネットや、イントラネットや、LAN（有線／無線の双方を含む）や、VAN や、パソコン通信網や、公衆電話網（アナログ／デジタルの双方を含む）や、専用回線網  
5 （アナログ／デジタルの双方を含む）や、CATV 網や、IMT2000 方式、GSM 方式または PDC／PDC-P 方式等の携帯回線交換網／携帯パケット交換網や、無線呼出網や、Bluetooth 等の局所無線網や、PHS 網や、CS、BS または ISDB 等の衛星通信網等のうちいずれかを含んでもよい。すなわち、本システムは、有線・無線を問わず任意のネットワークを介して、  
10 各種データを送受信することができる。

以上詳細に説明したように、本発明によれば、タンパク質の座標データを取得し、タンパク質の座標データについて、特定のアミノ酸残基から所定の距離内に含まれる近傍アミノ酸残基群の座標を抽出し、近傍アミノ酸残基群の切り口の部分にキャップ用置換基を付加し、キャップ用置換基が付加された近傍ア  
15 ミノ酸残基群の全体の電荷を計算し、キャップ用置換基が付加された近傍アミノ酸残基群について、計算された電荷を用いて特定のアミノ酸残基の原子座標について構造最適化を実行し、最適化された原子座標を、タンパク質の座標データ上の対応する原子座標と置換するので、水素位置の決定やパッキングの問題の解消を、実用的な計算資源を用いて行うことが可能となるタンパク質構造  
20 最適化装置、タンパク質構造最適化方法、プログラム、および、記録媒体を提供することができる。

また、本発明によれば、既存の計算プログラムには一切手を加えずに最適化処理の高速化を図ることができるタンパク質構造最適化装置、タンパク質構造最適化方法、プログラム、および、記録媒体を提供することができる。すなわ  
25 ち、既存の分子軌道計算プログラムや、分子力学計算プログラムの入出力ファイルを用いて、本装置を実行することができる。但し、本装置のアルゴリズムを既存の分子軌道計算プログラムや、分子力学計算プログラムに組み込むこと

も可能である。

また、本発明によれば、従来の方法では不可能な溶媒効果を考慮に入れたタンパク質の構造最適化が可能になるタンパク質構造最適化装置、タンパク質構造最適化方法、プログラム、および、記録媒体を提供することができる。

5       また、本発明によれば、キャップ用置換基は、水素原子（H）またはメチル基（CH<sub>3</sub>）であるので、近傍アミノ酸残基群について機械的に座標を切り取ったときの切り口が、ラジカルとなり計算に不都合を生じることを容易に解消することができるタンパク質構造最適化装置、タンパク質構造最適化方法、プログラム、および、記録媒体を提供することができる。

10       さらに、本発明によれば、抽出した近傍アミノ酸残基群の中にシステイン（CYS）が含まれている場合には、当該システイン（CYS）とジスルフィド結合をしておりかつ近傍アミノ酸残基群には含まれない別のシステイン（CYS）が存在しているか判定し、当該別のシステイン（CYS）が存在する場合  
15       には当該別のシステイン（CYS）も近傍アミノ酸残基群に加えるので、システイン間のジスルフィド結合を考慮して構造最適化を行うことができるタンパク質構造最適化装置、タンパク質構造最適化方法、プログラム、および、記録媒体を提供することができる。

#### 産業上の利用可能性

20       （I）以上のように、相互作用部位予測装置、相互作用部位予測方法、プログラム、および、記録媒体は、タンパク質の一次配列情報においてフラストレーションのある局所部位を発見することにより相互作用部位を効果的に予測することができる。

25       つまり、本発明にかかる相互作用部位予測装置、相互作用部位予測方法、プログラム、および、記録媒体は、特に、局所部位のフラストレーションに基づいて相互作用部位を予測することができる。

これにより、本発明にかかる相互作用部位予測装置、相互作用部位予測方法、

プログラム、および、記録媒体は、タンパク質の配列などの解析を行うバイオインフォマティクス分野において極めて有用である。また、本発明は、産業上多くの分野、特に医薬品、食品、化粧品、医療、遺伝子発現解析、タンパク質立体構造解析等の分野で広く実施することができ、極めて有用である。

- 5      (II) また、活性部位予測装置、活性部位予測方法、プログラム、および、記録媒体は、分子軌道計算によって得られた分子軌道のエネルギーや広がり的情報から蛋白質の活性部位を予測することができる。

つまり、本発明にかかる活性部位予測装置、活性部位予測方法、プログラム、および、記録媒体は、特に、生理活性ポリペプチドもしくは蛋白質の活性部位  
10      を高精度で推定することができる。

これにより、本発明にかかる活性部位予測装置、活性部位予測方法、プログラム、および、記録媒体は、タンパク質などの解析を行うバイオインフォマティクス分野において極めて有用である。また、本発明は、産業上多くの分野、特に医薬品、食品、化粧品、医療、遺伝子発現解析、タンパク質立体構造解析  
15      等の分野で広く実施することができ、極めて有用である。

(III) また、タンパク質相互作用情報処理装置、タンパク質相互作用情報処理方法、プログラム、および、記録媒体は、タンパク質の構造データから求めた疎水相互作用および静電相互作用に基づいてタンパク質単体時の不安定性の高い部位を特定し相互作用部位を特定すること等ができる。

20      これにより、本発明にかかるタンパク質相互作用情報処理装置、タンパク質相互作用情報処理方法、プログラム、および、記録媒体は、タンパク質などの解析を行うバイオインフォマティクス分野において極めて有用である。また、本発明は、産業上多くの分野、特に医薬品、食品、化粧品、医療、遺伝子発現解析、タンパク質立体構造解析等の分野で広く実施することができ、極めて有用である。  
25     

(IV) また、結合部位予測装置、結合部位予測方法、プログラム、および、記録媒体は、特に、アミノ酸配列データから予測された、または、実験的に求め

られた立体構造情報（アミノ酸残基間の空間上の距離情報）と、電荷の情報とを用いて静電的に不安定な部分を予測することにより、蛋白質または生理活性ポリペプチドの結合部位や結合相手を効率的に予測すること等ができる。

- つまり、本発明にかかる結合部位予測装置、結合部位予測方法、プログラム、  
5 および、記録媒体は、バイオインフォマティクスによる蛋白質の相互作用の予測を極めて短時間で計算可能にし、網羅的な解析を可能にする。

- これにより、本発明にかかる結合部位予測装置、結合部位予測方法、プログラム、および、記録媒体は、タンパク質などの解析を行うバイオインフォマティクス分野において極めて有用である。また、本発明は、産業上多くの分野、  
10 特に医薬品、食品、化粧品、医療、遺伝子発現解析、タンパク質立体構造解析等の分野で広く実施することができ、極めて有用である。

（V）また、タンパク質構造最適化装置、タンパク質構造最適化方法、プログラム、および、記録媒体は、蛋白質の構造を分割しながら所望の原子座標の最適化を行うことができる。

- 15 これにより、本発明にかかるタンパク質構造最適化装置、タンパク質構造最適化方法、プログラム、および、記録媒体は、タンパク質などの解析を行うバイオインフォマティクス分野において極めて有用である。また、本発明は、産業上多くの分野、特に医薬品、食品、化粧品、医療、遺伝子発現解析、タンパク質立体構造解析等の分野で広く実施することができ、極めて有用である。

## 請 求 の 範 囲

1. 目的のタンパク質の一次配列情報を入力する入力手段と、

タンパク質の一次配列情報から当該タンパク質の二次構造を予測する二次構造予測プログラムに対して上記入力手段により入力された上記一次配列情報の  
5 二次構造予測シミュレーションを実行させる二次構造予測プログラム実行手段と、

上記二次構造予測プログラム実行手段による上記二次構造予測プログラムの二次構造予測結果を比較する予測結果比較手段と、

10 上記予測結果比較手段による比較結果に基づいて、上記目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算するフラストレーション計算手段と、

上記フラストレーション計算手段により計算された上記局所部分の上記フラストレーションにより上記目的のタンパク質の相互作用部位を予測する相互作用部位予測手段と、  
15

を備えたことを特徴とする相互作用部位予測装置。

2. 目的のタンパク質の一次配列情報を入力する入力手段と、

上記目的のタンパク質の二次構造データを取得する二次構造データ取得手段  
20 と、

タンパク質の一次配列情報から当該タンパク質の二次構造を予測する二次構造予測プログラムに対して上記入力手段により入力された上記一次配列情報の二次構造予測シミュレーションを実行させる二次構造予測プログラム実行手段と、

25 上記二次構造予測プログラム実行手段による上記二次構造予測プログラムの二次構造予測結果と、上記二次構造データ取得手段により取得した上記二次構造データとを比較する予測結果比較手段と、



上記予測結果比較手段による比較結果に基づいて、上記目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算するフラストレーション計算手段と、

- 5   上記フラストレーション計算手段により計算された上記局所部分の上記フラストレーションにより上記目的のタンパク質の相互作用部位を予測する相互作用部位予測手段と、

を備えたことを特徴とする相互作用部位予測装置。

3.   上記二次構造予測プログラムの上記二次構造予測結果に対する確信度を  
10   示す確信度情報を設定する確信度情報設定手段、

をさらに備え、

- 上記フラストレーション計算手段は、上記確信度情報設定手段により設定された上記確信度情報および上記比較結果に基づいて、上記局所部分の上記フラストレーションを計算することを特徴とする請求の範囲第1項または第2項に  
15   記載の相互作用部位予測装置。

4.   目的のタンパク質の一次配列情報を入力する入力ステップと、

- タンパク質の一次配列情報から当該タンパク質の二次構造を予測する二次構造予測プログラムに対して上記入力ステップにより入力された上記一次配列情報の二次構造予測シミュレーションを実行させる二次構造予測プログラム実行  
20   ステップと、

上記二次構造予測プログラム実行ステップによる上記二次構造予測プログラムの二次構造予測結果を比較する予測結果比較ステップと、

- 上記予測結果比較ステップによる比較結果に基づいて、上記目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算するフラストレーション計算ステップと、  
25

上記フラストレーション計算ステップにより計算された上記局所部分の上記

フラストレーションにより上記目的のタンパク質の相互作用部位を予測する相互作用部位予測ステップと、

を含むことを特徴とする相互作用部位予測方法。

5      5.      目的のタンパク質の一次配列情報を入力する入力ステップと、

上記目的のタンパク質の二次構造データを取得する二次構造データ取得ステップと、

タンパク質の一次配列情報から当該タンパク質の二次構造を予測する二次構造予測プログラムに対して上記入力ステップにより入力された上記一次配列情報の二次構造予測シミュレーションを実行させる二次構造予測プログラム実行  
10      ステップと、

上記二次構造予測プログラム実行ステップによる上記二次構造予測プログラムの二次構造予測結果と、上記二次構造データ取得ステップにより取得した上記二次構造データとを比較する予測結果比較ステップと、

15      上記予測結果比較ステップによる比較結果に基づいて、上記目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算するフラストレーション計算ステップと、

上記フラストレーション計算ステップにより計算された上記局所部分の上記フラストレーションにより上記目的のタンパク質の相互作用部位を予測する相  
20      互作用部位予測ステップと、

を含むことを特徴とする相互作用部位予測方法。

6.      上記二次構造予測プログラムの上記二次構造予測結果に対する確信度を示す確信度情報を設定する確信度情報設定ステップ、

25      をさらに含み、

上記フラストレーション計算ステップは、上記確信度情報設定ステップにより設定された上記確信度情報および上記比較結果に基づいて、上記局所部分の

上記フラストレーションを計算することを特徴とする請求の範囲第4項または第5項に記載の相互作用部位予測方法。

7. 目的のタンパク質の一次配列情報を入力する入力ステップと、

- 5      タンパク質の一次配列情報から当該タンパク質の二次構造を予測する二次構造予測プログラムに対して上記入力ステップにより入力された上記一次配列情報の二次構造予測シミュレーションを実行させる二次構造予測プログラム実行ステップと、

- 10      上記二次構造予測プログラム実行ステップによる上記二次構造予測プログラムの二次構造予測結果を比較する予測結果比較ステップと、

上記予測結果比較ステップによる比較結果に基づいて、上記目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算するフラストレーション計算ステップと、

- 15      上記フラストレーション計算ステップにより計算された上記局所部分の上記フラストレーションにより上記目的のタンパク質の相互作用部位を予測する相互作用部位予測ステップと、

を含む相互作用部位予測方法をコンピュータに実行させることを特徴とするプログラム。

- 20      8. 目的のタンパク質の一次配列情報を入力する入力ステップと、

上記目的のタンパク質の二次構造データを取得する二次構造データ取得ステップと、

- 25      タンパク質の一次配列情報から当該タンパク質の二次構造を予測する二次構造予測プログラムに対して上記入力ステップにより入力された上記一次配列情報の二次構造予測シミュレーションを実行させる二次構造予測プログラム実行ステップと、

上記二次構造予測プログラム実行ステップによる上記二次構造予測プログラ

ムの二次構造予測結果と、上記二次構造データ取得ステップにより取得した上記二次構造データとを比較する予測結果比較ステップと、

上記予測結果比較ステップによる比較結果に基づいて、上記目的のタンパク質の一次配列情報の局所部分のフラストレーションを計算するフラストレーション計算ステップと、

上記フラストレーション計算ステップにより計算された上記局所部分の上記フラストレーションにより上記目的のタンパク質の相互作用部位を予測する相互作用部位予測ステップと、

を含む相互作用部位予測方法をコンピュータに実行させることを特徴とするプログラム。

9. 上記二次構造予測プログラムの上記二次構造予測結果に対する確信度を示す確信度情報を設定する確信度情報設定ステップ、

をさらに含み、

上記フラストレーション計算ステップは、上記確信度情報設定ステップにより設定された上記確信度情報および上記比較結果に基づいて、上記局所部分の上記フラストレーションを計算することを特徴とする請求の範囲第7項または第8項に記載のプログラム。

10. 上記請求の範囲第7項～第9項のいずれか一つに記載されたプログラムを記録したことを特徴とするコンピュータ読み取り可能な記録媒体。

11. 蛋白質または生理活性ポリペプチドの電子状態を分子軌道計算によって求めて、フロンティア軌道とその周辺軌道、および／または、主鎖の重原子に局在する軌道エネルギーを特定し、当該フロンティア軌道とその周辺軌道の位置、および／または、当該軌道エネルギーに基づいて当該蛋白質または当該生理活性ポリペプチドの活性部位となるアミノ酸残基を予測する活性部位予測

方法。

1 2. 目的の蛋白質または生理活性ポリペプチドの構造データを取得する構造データ取得ステップと、

5 上記構造データ取得ステップにて取得した上記構造データに基づいて上記蛋白質または上記生理活性ポリペプチドの電子状態を分子軌道計算によって求めてフロンティア軌道を特定するフロンティア軌道計算ステップと、

上記フロンティア軌道と予め定めたエネルギー差になる分子軌道を当該フロンティア軌道の周辺軌道として決定する周辺軌道決定ステップと、

10 上記フロンティア軌道と上記周辺軌道が分布しているアミノ酸残基を活性部位の候補のアミノ酸残基として決定する候補アミノ酸残基決定ステップと、

上記候補アミノ酸残基決定ステップにより決定された上記候補アミノ酸残基の中から活性部位を選択して活性部位を予測する活性部位予測ステップと、

を含むことを特徴とする活性部位予測方法。

15

1 3. 目的の蛋白質または生理活性ポリペプチドの構造データを取得する構造データ取得ステップと、

上記構造データ取得ステップにて取得した上記構造データに基づいて上記蛋白質または上記生理活性ポリペプチドの電子状態を分子軌道計算によって求め  
20 て主鎖の重原子に局在する軌道エネルギーを特定する軌道エネルギー計算ステップと、

上記軌道エネルギー計算ステップにて特定された上記軌道エネルギーのうち  
予め定めた値を超える軌道エネルギーとなる分子軌道および／または相対的に  
軌道エネルギーが高い分子軌道が分布しているアミノ酸残基を活性部位の候補  
25 のアミノ酸残基として決定する候補アミノ酸残基決定ステップと、

を含むことを特徴とする活性部位予測方法。

1 4. 目的の蛋白質または生理活性ポリペプチドの構造データを取得する構造データ取得ステップと、

上記構造データ取得ステップにて取得した上記構造データに基づいて上記蛋白質または上記生理活性ポリペプチドの電子状態を分子軌道計算によって求めてフロンティア軌道を特定するフロンティア軌道計算ステップと、

上記構造データ取得ステップにて取得した上記構造データに基づいて上記蛋白質または上記生理活性ポリペプチドの電子状態を分子軌道計算によって求めて主鎖の重原子に局在する軌道エネルギーを特定する軌道エネルギー計算ステップと、

10 上記フロンティア軌道と予め定めたエネルギー差になる分子軌道を当該フロンティア軌道の周辺軌道として決定する周辺軌道決定ステップと、

上記フロンティア軌道と上記周辺軌道が分布しているアミノ酸残基、および／または、上記軌道エネルギー計算ステップにて特定された上記軌道エネルギーのうち予め定めた値を超える軌道エネルギーとなる分子軌道および／または  
15 相対的に軌道エネルギーが高い分子軌道が分布しているアミノ酸残基を活性部位の候補のアミノ酸残基として決定する候補アミノ酸残基決定ステップと、

上記候補アミノ酸残基決定ステップにより決定された上記候補アミノ酸残基の中から活性部位を選択して活性部位を予測する活性部位予測ステップと、  
を含むことを特徴とする活性部位予測方法。

20

1 5. 分子軌道計算において、以下の3つの計算条件

- 1) 蛋白質または上記生理活性ポリペプチドのまわりに水分子を発生させる、
  - 2) 蛋白質または上記生理活性ポリペプチドの周りに連続的な誘電体を置く、
  - 3) 蛋白質または上記生理活性ポリペプチド表面の解離性アミノ酸残基を無  
25 電荷の状態にし、内部に埋まっている解離性アミノ酸を電荷状態にする、
- のうち少なくとも一つの計算条件を設定する計算条件設定ステップ、  
をさらに含むことを特徴とする請求の範囲第12項から第14項のいずれか

一つに記載の活性部位予測方法。

16. 目的の蛋白質または生理活性ポリペプチドの構造データを取得する構造データ取得手段と、

- 5 上記構造データ取得手段にて取得した上記構造データに基づいて上記蛋白質または上記生理活性ポリペプチドの電子状態を分子軌道計算によって求めてフロンティア軌道を特定するフロンティア軌道計算手段と、

上記フロンティア軌道と予め定めたエネルギー差になる分子軌道を当該フロンティア軌道の周辺軌道として決定する周辺軌道決定手段と、

- 10 上記フロンティア軌道と上記周辺軌道が分布しているアミノ酸残基を活性部位の候補のアミノ酸残基として決定する候補アミノ酸残基決定手段と、

上記候補アミノ酸残基決定手段により決定された上記候補アミノ酸残基の中から活性部位を選択して活性部位を予測する活性部位予測手段と、

を備えたことを特徴とする活性部位予測装置。

15

17. 目的の蛋白質または生理活性ポリペプチドの構造データを取得する構造データ取得手段と、

上記構造データ取得手段にて取得した上記構造データに基づいて上記蛋白質または上記生理活性ポリペプチドの電子状態を分子軌道計算によって求めて主鎖の重原子に局在する軌道エネルギーを特定する軌道エネルギー計算手段と、

- 20 上記軌道エネルギー計算手段にて特定された上記軌道エネルギーのうち予め定めた値を超える軌道エネルギーとなる分子軌道および／または相対的に軌道エネルギーが高い分子軌道が分布しているアミノ酸残基を活性部位の候補のアミノ酸残基として決定する候補アミノ酸残基決定手段と、

- 25 上記候補アミノ酸残基決定手段により決定された上記候補アミノ酸残基の中から活性部位を選択して活性部位を予測する活性部位予測手段と、

18. 目的の蛋白質または生理活性ポリペプチドの構造データを取得する構

造データ取得手段と、

上記構造データ取得手段にて取得した上記構造データに基づいて上記蛋白質または上記生理活性ポリペプチドの電子状態を分子軌道計算によって求めてフロンティア軌道を特定するフロンティア軌道計算手段と、

- 5      上記構造データ取得手段にて取得した上記構造データに基づいて上記蛋白質または上記生理活性ポリペプチドの電子状態を分子軌道計算によって求めて主鎖の重原子に局在する軌道エネルギーを特定する軌道エネルギー計算手段と、

上記フロンティア軌道と予め定めたエネルギー差になる分子軌道を当該フロンティア軌道の周辺軌道として決定する周辺軌道決定手段と、

- 10      上記フロンティア軌道と上記周辺軌道が分布しているアミノ酸残基、および／または、上記軌道エネルギー計算手段にて特定された上記軌道エネルギーのうち予め定めた値を超える軌道エネルギーとなる分子軌道および／または相対的に軌道エネルギーが高い分子軌道が分布しているアミノ酸残基を活性部位の候補のアミノ酸残基として決定する候補アミノ酸残基決定手段と、

- 15      上記候補アミノ酸残基決定手段により決定された上記候補アミノ酸残基の中から活性部位を選択して活性部位を予測する活性部位予測手段と、  
を備えたことを特徴とする活性部位予測装置。

#### 19. 分子軌道計算において、以下の3つの計算条件

- 20      1) 蛋白質または上記生理活性ポリペプチドのまわりに水分子を発生させる、  
2) 蛋白質または上記生理活性ポリペプチドの周りに連続的な誘電体を置く、  
3) 蛋白質または上記生理活性ポリペプチド表面の解離性アミノ酸残基を無電荷の状態にし、内部に埋まっている解離性アミノ酸を電荷状態にする、  
のうち少なくとも一つの計算条件を設定する計算条件設定手段、  
25      をさらに備えることを特徴とする請求の範囲第16項から第18項のいずれか一つに記載の活性部位予測装置。



20. 目的の蛋白質または生理活性ポリペプチドの構造データを取得する構造データ取得ステップと、

上記構造データ取得ステップにて取得した上記構造データに基づいて上記蛋白質または上記生理活性ポリペプチドの電子状態を分子軌道計算によって求めてフロンティア軌道を特定するフロンティア軌道計算ステップと、

上記フロンティア軌道と予め定めたエネルギー差になる分子軌道を当該フロンティア軌道の周辺軌道として決定する周辺軌道決定ステップと、

上記フロンティア軌道と上記周辺軌道が分布しているアミノ酸残基を活性部位の候補のアミノ酸残基として決定する候補アミノ酸残基決定ステップと、

上記候補アミノ酸残基決定ステップにより決定された上記候補アミノ酸残基の中から活性部位を選択して活性部位を予測する活性部位予測ステップと、

を含む活性部位予測方法をコンピュータに実行させることを特徴とするプログラム。

21. 目的の蛋白質または生理活性ポリペプチドの構造データを取得する構造データ取得ステップと、

上記構造データ取得ステップにて取得した上記構造データに基づいて上記蛋白質または上記生理活性ポリペプチドの電子状態を分子軌道計算によって求めて主鎖の重原子に局在する軌道エネルギーを特定する軌道エネルギー計算ステップと、

上記軌道エネルギー計算ステップにて特定された上記軌道エネルギーのうち予め定めた値を超える軌道エネルギーとなる分子軌道および／または相対的に軌道エネルギーが高い分子軌道が分布しているアミノ酸残基を活性部位の候補のアミノ酸残基として決定する候補アミノ酸残基決定ステップと、

を含む活性部位予測方法をコンピュータに実行させることを特徴とするプログラム。

22. 目的の蛋白質または生理活性ポリペプチドの構造データを取得する構造データ取得ステップと、

上記構造データ取得ステップにて取得した上記構造データに基づいて上記蛋白質または上記生理活性ポリペプチドの電子状態を分子軌道計算によって求めてフロンティア軌道を特定するフロンティア軌道計算ステップと、

上記構造データ取得ステップにて取得した上記構造データに基づいて上記蛋白質または上記生理活性ポリペプチドの電子状態を分子軌道計算によって求めて主鎖の重原子に局在する軌道エネルギーを特定する軌道エネルギー計算ステップと、

上記フロンティア軌道と予め定めたエネルギー差になる分子軌道を当該フロンティア軌道の周辺軌道として決定する周辺軌道決定ステップと、

上記フロンティア軌道と上記周辺軌道が分布しているアミノ酸残基、および／または、上記軌道エネルギー計算ステップにて特定された上記軌道エネルギーのうち予め定めた値を超える軌道エネルギーとなる分子軌道および／または相対的に軌道エネルギーが高い分子軌道が分布しているアミノ酸残基を活性部位の候補のアミノ酸残基として決定する候補アミノ酸残基決定ステップと、

上記候補アミノ酸残基決定ステップにより決定された上記候補アミノ酸残基の中から活性部位を選択して活性部位を予測する活性部位予測ステップと、

を含む活性部位予測方法をコンピュータに実行させることを特徴とするプログラム。

23. 分子軌道計算において、以下の3つの計算条件

- 1) 蛋白質または上記生理活性ポリペプチドのまわりに水分子を発生させる、
  - 2) 蛋白質または上記生理活性ポリペプチドの周りに連続的な誘電体を置く、
  - 3) 蛋白質または上記生理活性ポリペプチド表面の解離性アミノ酸残基を無電荷の状態にし、内部に埋まっている解離性アミノ酸を電荷状態にする、
- のうち少なくとも一つの計算条件を設定する計算条件設定ステップ、

をさらに含む活性部位予測方法をコンピュータに実行させることを特徴とする請求の範囲第20項から第22項のいずれか一つに記載のプログラム。

24. 上記請求の範囲第20項から第23項のいずれか一つに記載されたプログラムを記録したことを特徴とするコンピュータ読み取り可能な記録媒体。

25. 相互作用する複数のタンパク質の一次構造データと単体時および／または複合体形成時の立体構造データとを含む構造データを取得する構造データ取得手段と、  
10 上記構造データ取得手段にて取得した上記構造データに従って、上記一次構造データを構成するアミノ酸残基毎に疎水性相互作用エネルギーを特定する疎水面特定手段と、

上記構造データ取得手段にて取得した上記構造データに従って、上記一次構造データを構成するアミノ酸残基毎に静電相互作用エネルギーを特定する静電相互作用部位特定手段と、  
15

上記疎水面特定手段にて特定された上記疎水性相互作用エネルギー、および、上記静電相互作用部位特定手段にて特定された上記静電相互作用エネルギーに基づいて不安定性の高い上記アミノ酸残基の部位を特定することにより相互作用部位を特定する相互作用部位特定手段と、  
20 上記を備えたことを特徴とするタンパク質相互作用情報処理装置。

26. 上記構造データ取得手段にて取得した上記構造データに従って、上記一次構造データを構成するアミノ酸残基毎に溶媒接触面を特定する溶媒接触面特定手段、  
25 上記をさらに備え、

上記相互作用部位特定手段は、上記溶媒接触面特定手段にて特定された上記溶媒接触面、上記疎水面特定手段にて特定された上記疎水性相互作用エネルギー

一、および、上記静電相互作用部位特定手段にて特定された上記静電相互作用エネルギーに基づいて不安定性の高い上記アミノ酸残基の部位を特定することにより相互作用部位を特定することを特徴とする請求の範囲第25項に記載のタンパク質相互作用情報処理装置。

5

27. 上記相互作用部位特定手段にて特定された上記相互作用部位について、相互作用する相手側の一次配列を特定し、当該一次配列を含む一次構造を持つ候補タンパク質を検索する候補タンパク質検索手段、

をさらに備え、

10

上記候補タンパク質検索手段にて検索された上記候補タンパク質について、上記相手側の一次配列の部分が上記候補タンパク質の相互作用部位として特定されるか否かを確認することを特徴とする請求の範囲第25項または第26項に記載のタンパク質相互作用情報処理装置。

15

28. 相互作用する複数のタンパク質の一次構造データと単体時および／または複合体形成時の立体構造データとを含む構造データを取得する構造データ取得ステップと、

上記構造データ取得ステップにて取得した上記構造データに従って、上記一次構造データを構成するアミノ酸残基毎に疎水性相互作用エネルギーを特定する疎水面特定ステップと、

20

上記構造データ取得ステップにて取得した上記構造データに従って、上記一次構造データを構成するアミノ酸残基毎に静電相互作用エネルギーを特定する静電相互作用部位特定ステップと、

上記疎水面特定ステップにて特定された上記疎水性相互作用エネルギー、および、上記静電相互作用部位特定ステップにて特定された上記静電相互作用エネルギーに基づいて不安定性の高い上記アミノ酸残基の部位を特定することにより相互作用部位を特定する相互作用部位特定ステップと、

25

を含むことを特徴とするタンパク質相互作用情報処理方法。

29. 上記構造データ取得ステップにて取得した上記構造データに従って、  
上記一次構造データを構成するアミノ酸残基毎に溶媒接触面を特定する溶媒接  
5 触面特定ステップ、

をさらに含み、

上記相互作用部位特定ステップは、上記溶媒接触面特定ステップにて特定さ  
れた上記溶媒接触面、上記疎水面特定ステップにて特定された上記疎水性相互  
作用エネルギー、および、上記静電相互作用部位特定ステップにて特定された  
10 上記静電相互作用エネルギーに基づいて不安定性の高い上記アミノ酸残基の部  
位を特定することにより相互作用部位を特定することを特徴とする請求の範囲  
第28項に記載のタンパク質相互作用情報処理方法。

30. 上記相互作用部位特定ステップにて特定された上記相互作用部位につ  
15 いて、相互作用する相手側の一次配列を特定し、当該一次配列を含む一次構造  
を持つ候補タンパク質を検索する候補タンパク質検索ステップ、

をさらに含み、

上記候補タンパク質検索ステップにて検索された上記候補タンパク質につい  
て、上記相手側の一次配列の部分が上記候補タンパク質の相互作用部位として  
20 特定されるか否かを確認することを特徴とする請求の範囲第28項または第2  
9項に記載のタンパク質相互作用情報処理方法。

31. 相互作用する複数のタンパク質の一次構造データと単体時および／ま  
たは複合体形成時の立体構造データとを含む構造データを取得する構造データ  
25 取得ステップと、

上記構造データ取得ステップにて取得した上記構造データに従って、上記一  
次構造データを構成するアミノ酸残基毎に疎水性相互作用エネルギーを特定す

る疎水面特定ステップと、

上記構造データ取得ステップにて取得した上記構造データに従って、上記一次構造データを構成するアミノ酸残基毎に静電相互作用エネルギーを特定する静電相互作用部位特定ステップと、

- 5      上記疎水面特定ステップにて特定された上記疎水性相互作用エネルギー、および、上記静電相互作用部位特定ステップにて特定された上記静電相互作用エネルギーに基づいて不安定性の高い上記アミノ酸残基の部位を特定することにより相互作用部位を特定する相互作用部位特定ステップと、

- 10      を含むタンパク質相互作用情報処理方法をコンピュータに実行させることを特徴とするプログラム。

3 2.   上記構造データ取得ステップにて取得した上記構造データに従って、上記一次構造データを構成するアミノ酸残基毎に溶媒接触面を特定する溶媒接触面特定ステップ、

- 15      をさらに含み、

- 上記相互作用部位特定ステップは、上記溶媒接触面特定ステップにて特定された上記溶媒接触面、上記疎水面特定ステップにて特定された上記疎水性相互作用エネルギー、および、上記静電相互作用部位特定ステップにて特定された上記静電相互作用エネルギーに基づいて不安定性の高い上記アミノ酸残基の部位を特定することにより相互作用部位を特定することを特徴とする請求の範囲第31項に記載のプログラム。
- 20

- 3 3.   上記相互作用部位特定ステップにて特定された上記相互作用部位について、相互作用する相手側の一次配列を特定し、当該一次配列を含む一次構造を持つ候補タンパク質を検索する候補タンパク質検索ステップ、
- 25

    をさらに含み、

    上記候補タンパク質検索ステップにて検索された上記候補タンパク質につい

て、上記相手側の一次配列の部分が上記候補タンパク質の相互作用部位として特定されるか否かを確認することを特徴とする請求の範囲第31項または第32項に記載のプログラム。

- 5     34.     上記請求の範囲第31項から第33項のいずれか一つに記載されたプログラムを記録したことを特徴とするコンピュータ読み取り可能な記録媒体。

35.     蛋白質または生理活性ポリペプチドのアミノ酸配列データから当該蛋白質または生理活性ポリペプチドの立体構造における各アミノ酸残基間の空間上の距離データを求め、当該距離データと各アミノ酸の電荷に従って静電的に不安定なアミノ酸残基を特定することにより結合部位を予測することを特徴とする結合部位予測方法。

10

36.     目的の蛋白質または生理活性ポリペプチドのアミノ酸配列データを取得するアミノ酸配列データ取得ステップと、

15

上記アミノ酸配列データ取得ステップにより取得されたアミノ酸配列データに含まれる各アミノ酸残基の間の空間上の距離を決定する空間距離決定ステップと、

上記アミノ酸配列データに含まれる各アミノ酸残基が持つ電荷を決定する電荷決定ステップと、

20

上記空間距離決定ステップにより決定された各アミノ酸残基間の空間上の距離と、上記電荷決定ステップにより決定された各アミノ酸残基が持つ電荷とに従って、各アミノ酸残基のエネルギーを計算するエネルギー計算ステップと、

上記エネルギー計算ステップにより計算された上記エネルギーに従って、結合部位となる候補アミノ酸残基を決定する候補アミノ酸残基決定ステップと、

25

を含むことを特徴とする結合部位予測方法。

37. 目的の複数の蛋白質または生理活性ポリペプチドのアミノ酸配列データを取得するアミノ酸配列データ取得ステップと、

上記目的の複数の蛋白質または生理活性ポリペプチドの結合した複合体の立体構造情報を生成する複合体構造生成ステップと、

- 5      上記アミノ酸配列データ取得ステップにより取得された複数のアミノ酸配列データに含まれる各アミノ酸残基の間の空間上の距離を、上記複合体構造生成ステップにより生成した上記複合体の立体構造情報に従って決定する空間距離決定ステップと、

- 10      上記複数のアミノ酸配列データに含まれる各アミノ酸残基が持つ電荷を決定する電荷決定ステップと、

上記空間距離決定ステップにより決定された各アミノ酸残基間の空間上の距離と、上記電荷決定ステップにより決定された各アミノ酸残基が持つ電荷とに従って、各アミノ酸残基のエネルギーを計算するエネルギー計算ステップと、

- 15      上記複合体構造生成ステップにより上記複合体について結合部位を変えて上記複合体の立体構造情報を生成し、上記エネルギー計算ステップにより各アミノ酸残基のエネルギーを計算し、当該エネルギーの総和が最小となる結合部位を求めるエネルギー最小化ステップと、

- 20      上記エネルギー最小化ステップによりエネルギーの総和が最小となる結合部位を、結合部位の候補アミノ酸残基として決定する候補アミノ酸残基決定ステップと、

を含むことを特徴とする結合部位予測方法。

38. 目的の蛋白質または生理活性ポリペプチドのアミノ酸配列データと、結合候補となる1つまたは複数の蛋白質または生理活性ポリペプチドのアミノ酸配列データとを取得するアミノ酸配列データ取得ステップと、
- 25

上記目的の蛋白質または生理活性ポリペプチドと、上記結合候補となる蛋白質または生理活性ポリペプチドとが結合した複合体の立体構造情報を生成する



複合体構造生成ステップと、

上記アミノ酸配列データ取得ステップにより取得された目的のアミノ酸配列データと結合候補のアミノ酸配列データに含まれる各アミノ酸残基の間の空間上の距離を、上記複合体構造生成ステップにより生成した上記複合体の立体構造情報に従って決定する空間距離決定ステップと、

上記目的のアミノ酸配列データと上記結合候補のアミノ酸配列データに含まれる各アミノ酸残基が持つ電荷を決定する電荷決定ステップと、

上記空間距離決定ステップにより決定された各アミノ酸残基間の空間上の距離と、上記電荷決定ステップにより決定された各アミノ酸残基が持つ電荷とに従って、各アミノ酸残基のエネルギーを計算するエネルギー計算ステップと、

上記複合体構造生成ステップにより上記複合体について結合部位を変えて上記複合体の立体構造情報を生成し、上記エネルギー計算ステップにより各アミノ酸残基のエネルギーを計算し、当該エネルギーの総和が最小となる結合部位を求めるエネルギー最小化ステップと、

全ての結合候補について上記エネルギー最小化ステップを実行した結果、上記エネルギーの総和が最小となる結合部位をもつ結合候補を決定する結合候補決定ステップと、

を含むことを特徴とする結合部位予測方法。

39. 目的の蛋白質または生理活性ポリペプチドのアミノ酸配列データを取得するアミノ酸配列データ取得手段と、

上記アミノ酸配列データ取得手段により取得されたアミノ酸配列データに含まれる各アミノ酸残基の間の空間上の距離を決定する空間距離決定手段と、

上記アミノ酸配列データに含まれる各アミノ酸残基が持つ電荷を決定する電荷決定手段と、

上記空間距離決定手段により決定された各アミノ酸残基間の空間上の距離と、上記電荷決定手段により決定された各アミノ酸残基が持つ電荷とに従って、各

アミノ酸残基のエネルギーを計算するエネルギー計算手段と、

上記エネルギー計算手段により計算された上記エネルギーに従って、結合部位となる候補アミノ酸残基を決定する候補アミノ酸残基決定手段と、  
を備えたことを特徴とする結合部位予測装置。

5

40. 目的の複数の蛋白質または生理活性ポリペプチドのアミノ酸配列データを取得するアミノ酸配列データ取得手段と、

上記目的の複数の蛋白質または生理活性ポリペプチドの結合した複合体の立体構造情報を生成する複合体構造生成手段と、

10 上記アミノ酸配列データ取得手段により取得された複数のアミノ酸配列データに含まれる各アミノ酸残基の間の空間上の距離を、上記複合体構造生成手段により生成した上記複合体の立体構造情報に従って決定する空間距離決定手段と、

上記複数のアミノ酸配列データに含まれる各アミノ酸残基が持つ電荷を決定  
15 する電荷決定手段と、

上記空間距離決定手段により決定された各アミノ酸残基間の空間上の距離と、  
上記電荷決定手段により決定された各アミノ酸残基が持つ電荷とに従って、各  
アミノ酸残基のエネルギーを計算するエネルギー計算手段と、

上記複合体構造生成手段により上記複合体について結合部位を変えて上記複  
20 合体の立体構造情報を生成し、上記エネルギー計算手段により各アミノ酸残基  
のエネルギーを計算し、当該エネルギーの総和が最小となる結合部位を求める  
エネルギー最小化手段と、

上記エネルギー最小化手段によりエネルギーの総和が最小となる結合部位を、  
結合部位の候補アミノ酸残基として決定する候補アミノ酸残基決定手段と、

25 を備えたことを特徴とする結合部位予測装置。

41. 目的の蛋白質または生理活性ポリペプチドのアミノ酸配列データと、

結合候補となる1つまたは複数の蛋白質または生理活性ポリペプチドのアミノ酸配列データとを取得するアミノ酸配列データ取得手段と、

- 上記目的の蛋白質または生理活性ポリペプチドと、上記結合候補となる蛋白質または生理活性ポリペプチドとが結合した複合体の立体構造情報を生成する  
5 複合体構造生成手段と、

上記アミノ酸配列データ取得手段により取得された目的のアミノ酸配列データと結合候補のアミノ酸配列データに含まれる各アミノ酸残基の間の空間上の距離を、上記複合体構造生成手段により生成した上記複合体の立体構造情報に従って決定する空間距離決定手段と、

- 10 上記目的のアミノ酸配列データと上記結合候補のアミノ酸配列データに含まれる各アミノ酸残基が持つ電荷を決定する電荷決定手段と、

上記空間距離決定手段により決定された各アミノ酸残基間の空間上の距離と、上記電荷決定手段により決定された各アミノ酸残基が持つ電荷とに従って、各アミノ酸残基のエネルギーを計算するエネルギー計算手段と、

- 15 上記複合体構造生成手段により上記複合体について結合部位を変えて上記複合体の立体構造情報を生成し、上記エネルギー計算手段により各アミノ酸残基のエネルギーを計算し、当該エネルギーの総和が最小となる結合部位を求めるエネルギー最小化手段と、

- 全ての結合候補について上記エネルギー最小化手段を実行した結果、上記エ  
20 ネルギーの総和が最小となる結合部位をもつ結合候補を決定する結合候補決定手段と、

を備えたことを特徴とする結合部位予測装置。

- 4 2. 目的の蛋白質または生理活性ポリペプチドのアミノ酸配列データを取  
25 得するアミノ酸配列データ取得ステップと、

上記アミノ酸配列データ取得ステップにより取得されたアミノ酸配列データに含まれる各アミノ酸残基の間の空間上の距離を決定する空間距離決定ステッ

ブと、

上記アミノ酸配列データに含まれる各アミノ酸残基が持つ電荷を決定する電荷決定ステップと、

5 上記空間距離決定ステップにより決定された各アミノ酸残基間の空間上の距離と、上記電荷決定ステップにより決定された各アミノ酸残基が持つ電荷とに従って、各アミノ酸残基のエネルギーを計算するエネルギー計算ステップと、

上記エネルギー計算ステップにより計算された上記エネルギーに従って、結合部位となる候補アミノ酸残基を決定する候補アミノ酸残基決定ステップと、

10 を含む結合部位予測方法をコンピュータに実行させることを特徴とするプログラム。

43. 目的の複数の蛋白質または生理活性ポリペプチドのアミノ酸配列データを取得するアミノ酸配列データ取得ステップと、

15 上記目的の複数の蛋白質または生理活性ポリペプチドの結合した複合体の立体構造情報を生成する複合体構造生成ステップと、

上記アミノ酸配列データ取得ステップにより取得された複数のアミノ酸配列データに含まれる各アミノ酸残基の間の空間上の距離を、上記複合体構造生成ステップにより生成した上記複合体の立体構造情報に従って決定する空間距離決定ステップと、

20 上記複数のアミノ酸配列データに含まれる各アミノ酸残基が持つ電荷を決定する電荷決定ステップと、

上記空間距離決定ステップにより決定された各アミノ酸残基間の空間上の距離と、上記電荷決定ステップにより決定された各アミノ酸残基が持つ電荷とに従って、各アミノ酸残基のエネルギーを計算するエネルギー計算ステップと、

25 上記複合体構造生成ステップにより上記複合体について結合部位を変えて上記複合体の立体構造情報を生成し、上記エネルギー計算ステップにより各アミノ酸残基のエネルギーを計算し、当該エネルギーの総和が最小となる結合部位

を求めるエネルギー最小化ステップと、

上記エネルギー最小化ステップによりエネルギーの総和が最小となる結合部位を、結合部位の候補アミノ酸残基として決定する候補アミノ酸残基決定ステップと、

- 5       を含む結合部位予測方法をコンピュータに実行させることを特徴とするプログラム。

- 4 4.    目的の蛋白質または生理活性ポリペプチドのアミノ酸配列データと、結合候補となる1つまたは複数の蛋白質または生理活性ポリペプチドのアミノ酸配列データとを取得するアミノ酸配列データ取得ステップと、

上記目的の蛋白質または生理活性ポリペプチドと、上記結合候補となる蛋白質または生理活性ポリペプチドとが結合した複合体の立体構造情報を生成する複合体構造生成ステップと、

- 15       上記アミノ酸配列データ取得ステップにより取得された目的のアミノ酸配列データと結合候補のアミノ酸配列データに含まれる各アミノ酸残基の間の空間上の距離を、上記複合体構造生成ステップにより生成した上記複合体の立体構造情報に従って決定する空間距離決定ステップと、

上記目的のアミノ酸配列データと上記結合候補のアミノ酸配列データに含まれる各アミノ酸残基が持つ電荷を決定する電荷決定ステップと、

- 20       上記空間距離決定ステップにより決定された各アミノ酸残基間の空間上の距離と、上記電荷決定ステップにより決定された各アミノ酸残基が持つ電荷とに従って、各アミノ酸残基のエネルギーを計算するエネルギー計算ステップと、

- 上記複合体構造生成ステップにより上記複合体について結合部位を変えて上記複合体の立体構造情報を生成し、上記エネルギー計算ステップにより各アミノ酸残基のエネルギーを計算し、当該エネルギーの総和が最小となる結合部位を求めるエネルギー最小化ステップと、

全ての結合候補について上記エネルギー最小化ステップを実行した結果、上

記エネルギーの総和が最小となる結合部位をもつ結合候補を決定する結合候補決定ステップと、

を含む結合部位予測方法をコンピュータに実行させることを特徴とするプログラム。

5

45. 上記請求の範囲第42項から第44項のいずれか一つに記載されたプログラムを記録したことを特徴とするコンピュータ読み取り可能な記録媒体。

46. タンパク質の座標データを取得する座標データ取得手段と、

10 上記タンパク質の座標データについて、特定のアミノ酸残基から所定の距離内に含まれる近傍アミノ酸残基群の座標を抽出する近傍アミノ酸残基群抽出手段と、

上記近傍アミノ酸残基群の切り口の部分にキャップ用置換基を付加するキャップ付加手段と、

15 上記キャップ付加手段により上記キャップ用置換基が付加された上記近傍アミノ酸残基群の全体の電荷を計算する電荷計算手段と、

上記キャップ付加手段により上記キャップ用置換基が付加された上記近傍アミノ酸残基群について、上記電荷計算手段により計算された上記電荷を用いて上記特定のアミノ酸残基の原子座標について構造最適化を実行する構造最適化  
20 手段と、

上記構造最適化手段にて最適化された上記原子座標を、上記タンパク質の座標データ上の対応する原子座標と置換する原子座標置換手段と、

を備えたことを特徴とするタンパク質構造最適化装置。

25 47. 上記キャップ用置換基は、水素原子(H)またはメチル基(CH<sub>3</sub>)であること、

を特徴とする請求の範囲第46項に記載のタンパク質構造最適化装置。

4 8. 上記近傍アミノ酸残基群抽出手段は、

抽出した上記近傍アミノ酸残基群の中にシステイン（C Y S）が含まれている場合には、当該システイン（C Y S）とジスルフィド結合をしておりかつ上記近傍アミノ酸残基群には含まれない別のシステイン（C Y S）が存在しているか判定し、当該別のシステイン（C Y S）が存在する場合には当該別のシステイン（C Y S）も近傍アミノ酸残基群に加えること、

を特徴とする請求の範囲第46項または第47項に記載のタンパク質構造最適化装置。

10

4 9. タンパク質の座標データを取得する座標データ取得ステップと、

上記タンパク質の座標データについて、特定のアミノ酸残基から所定の距離内に含まれる近傍アミノ酸残基群の座標を抽出する近傍アミノ酸残基群抽出ステップと、

15 上記近傍アミノ酸残基群の切り口の部分にキャップ用置換基を付加するキャップ付加ステップと、

上記キャップ付加ステップにより上記キャップ用置換基が付加された上記近傍アミノ酸残基群の全体の電荷を計算する電荷計算ステップと、

20 上記キャップ付加ステップにより上記キャップ用置換基が付加された上記近傍アミノ酸残基群について、上記電荷計算ステップにより計算された上記電荷を用いて上記特定のアミノ酸残基の原子座標について構造最適化を実行する構造最適化ステップと、

上記構造最適化ステップにて最適化された上記原子座標を、上記タンパク質の座標データ上の対応する原子座標と置換する原子座標置換ステップと、

25 を含むことを特徴とするタンパク質構造最適化方法。

50. 上記キャップ用置換基は、水素原子 (H) またはメチル基 ( $\text{CH}_3$ ) であること、

を特徴とする請求の範囲第49項に記載のタンパク質構造最適化方法。

5 51. 上記近傍アミノ酸残基群抽出ステップは、

抽出した上記近傍アミノ酸残基群の中にシステイン (CYS) が含まれている場合には、当該システイン (CYS) とジスルフィド結合をしておりかつ上記近傍アミノ酸残基群には含まれない別のシステイン (CYS) が存在しているか判定し、当該別のシステイン (CYS) が存在する場合には当該別のシステイン (CYS) も近傍アミノ酸残基群に加えること、

10 を特徴とする請求の範囲第49項または第50項に記載のタンパク質構造最適化方法。

52. タンパク質の座標データを取得する座標データ取得ステップと、

15 上記タンパク質の座標データについて、特定のアミノ酸残基から所定の距離内に含まれる近傍アミノ酸残基群の座標を抽出する近傍アミノ酸残基群抽出ステップと、

上記近傍アミノ酸残基群の切り口の部分にキャップ用置換基を付加するキャップ付加ステップと、

20 上記キャップ付加ステップにより上記キャップ用置換基が付加された上記近傍アミノ酸残基群の全体の電荷を計算する電荷計算ステップと、

上記キャップ付加ステップにより上記キャップ用置換基が付加された上記近傍アミノ酸残基群について、上記電荷計算ステップにより計算された上記電荷を用いて上記特定のアミノ酸残基の原子座標について構造最適化を実行する構造最適化ステップと、

25 上記構造最適化ステップにて最適化された上記原子座標を、上記タンパク質の座標データ上の対応する原子座標と置換する原子座標置換ステップと、



を含むタンパク質構造最適化方法をコンピュータに実行させることを特徴とするプログラム。

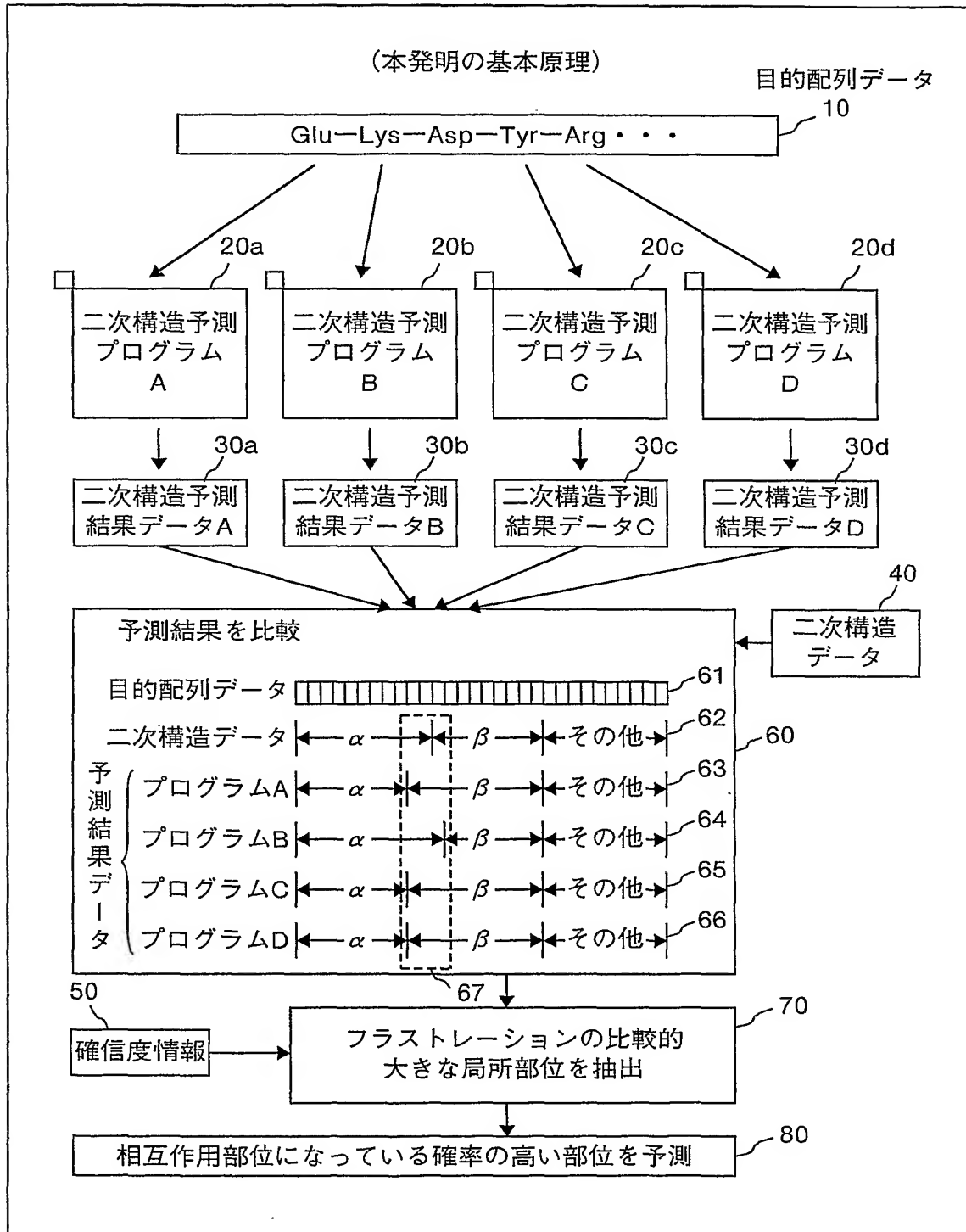
- 5 3. 上記キャップ用置換基は、水素原子 (H) またはメチル基 ( $\text{CH}_3$ ) であること、  
5 を特徴とする請求の範囲第 5 2 項に記載のプログラム。

- 5 4. 上記近傍アミノ酸残基群抽出ステップは、  
抽出した上記近傍アミノ酸残基群の中にシステイン (CYS) が含まれている場合には、当該システイン (CYS) とジスルフィド結合をしておりかつ上記近傍アミノ酸残基群には含まれない別のシステイン (CYS) が存在しているか判定し、当該別のシステイン (CYS) が存在する場合には当該別のシステイン (CYS) も近傍アミノ酸残基群に加えること、  
10 を特徴とする請求の範囲第 5 2 項または第 5 3 項に記載のプログラム。

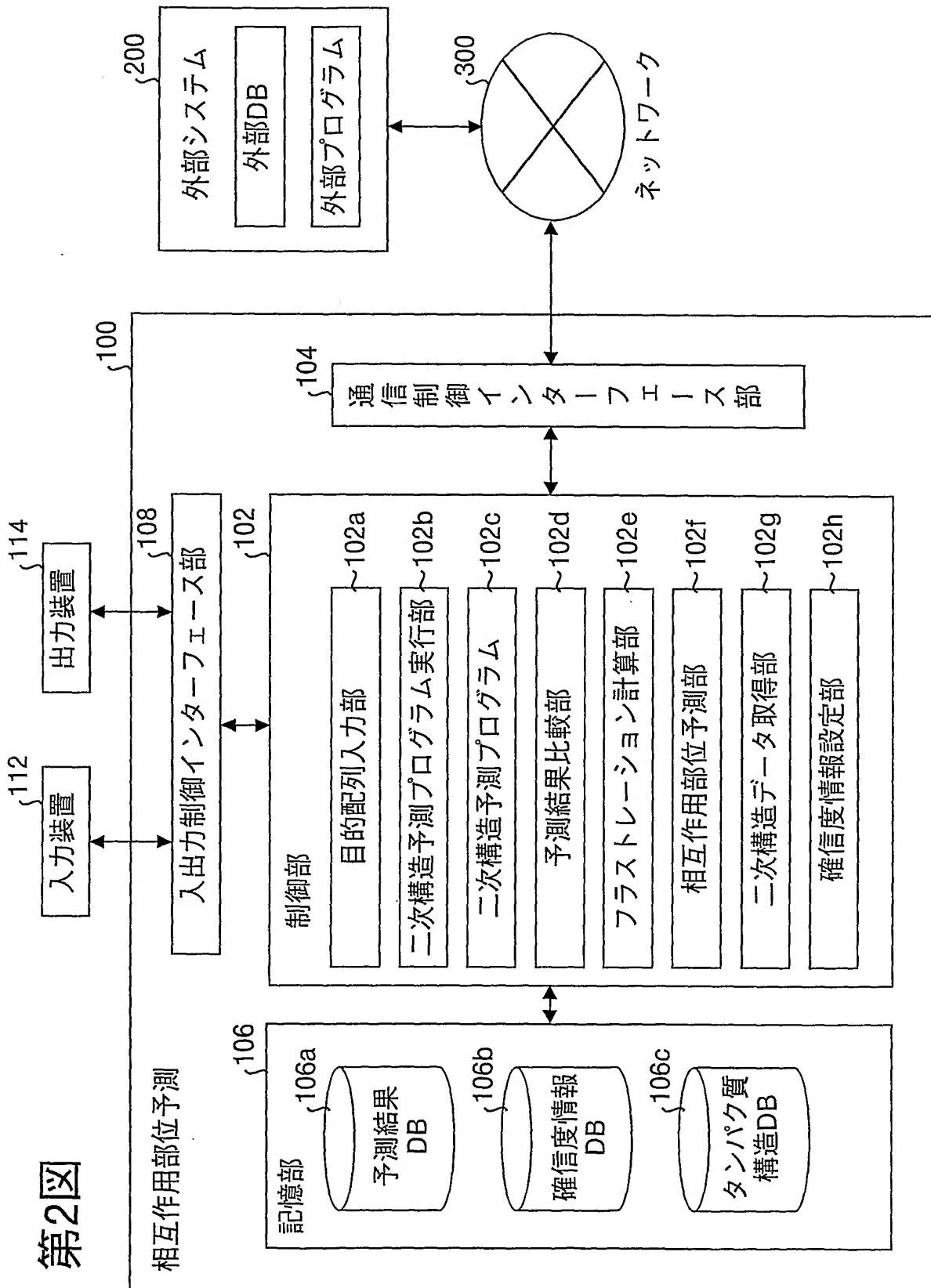
15

- 5 5. 上記請求の範囲第 5 2 項から 5 4 項のいずれか一つに記載されたプログラムを記録したことを特徴とするコンピュータ読み取り可能な記録媒体。

## 第1図



2/77

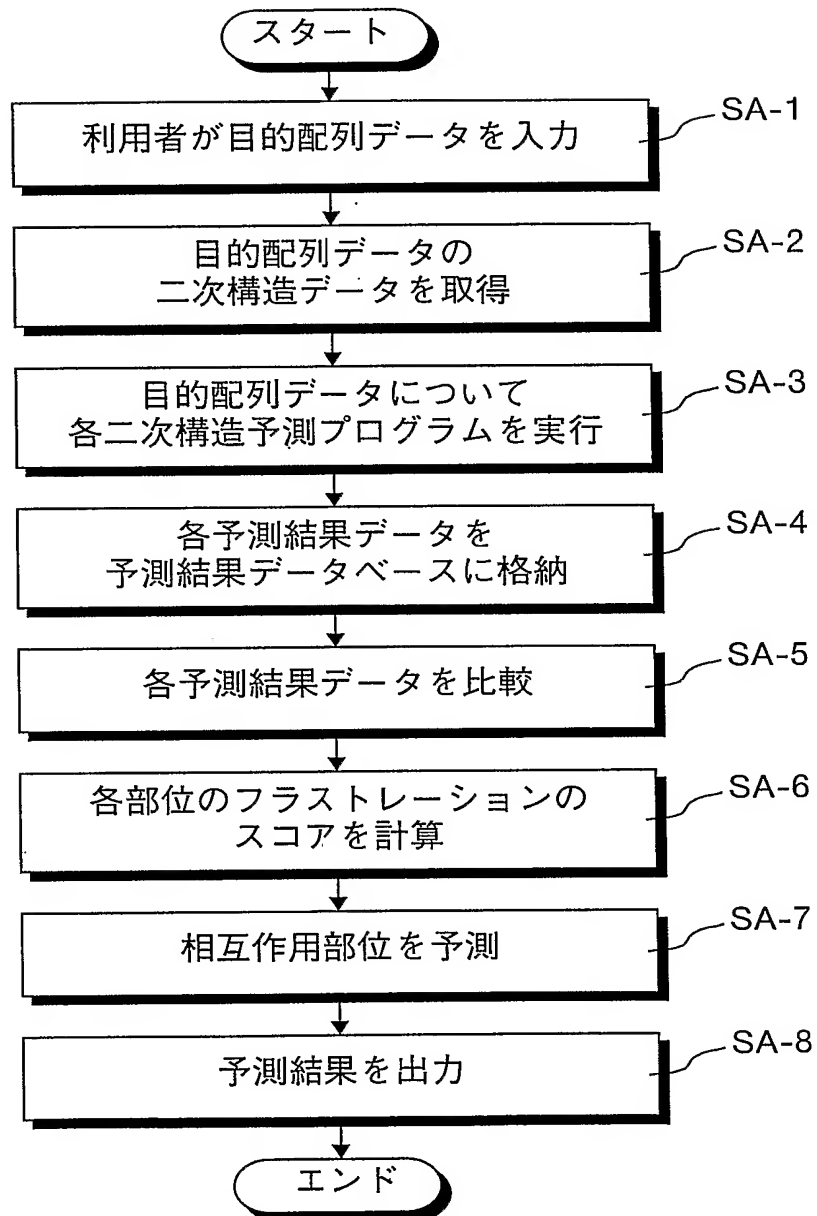


第3図

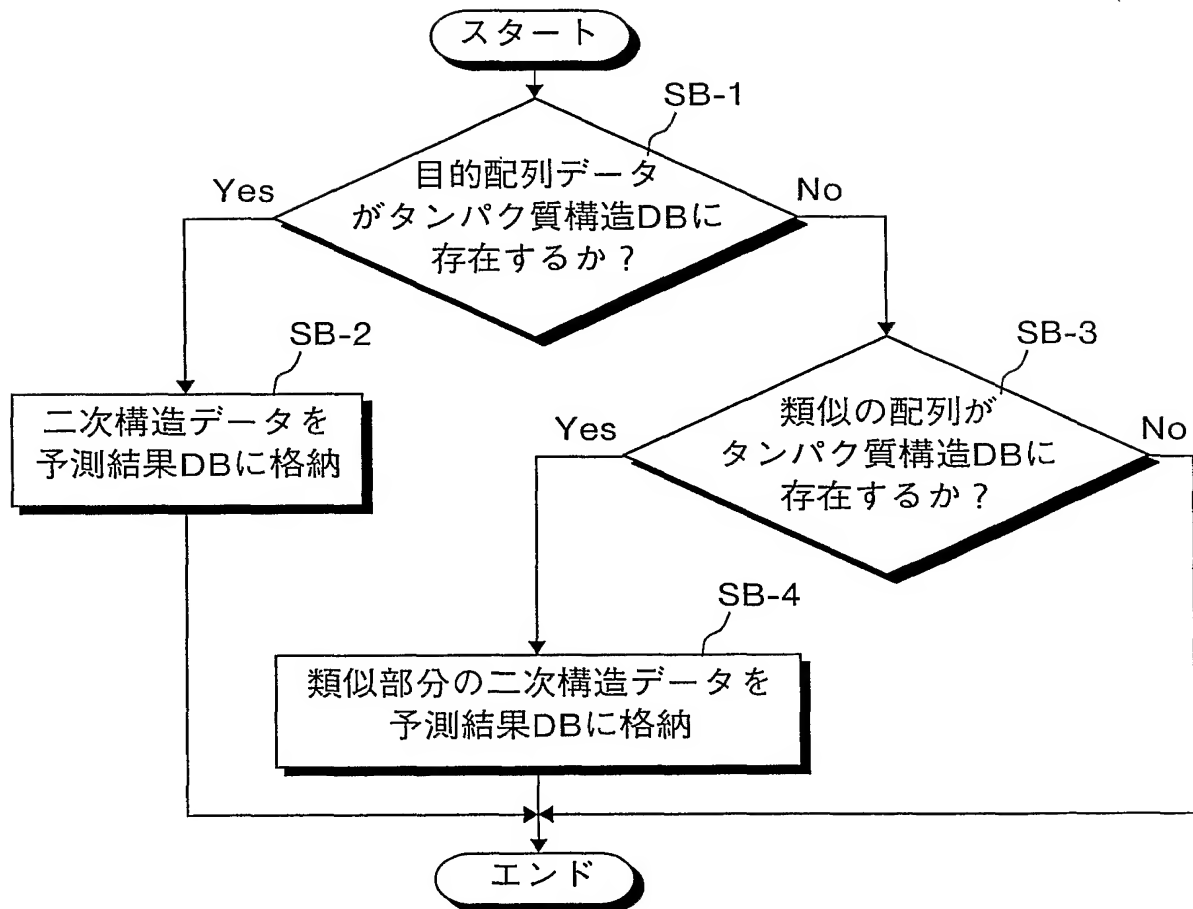
106a 予測結果データベース

目的配列 データ	二次構造 データ	予測結果データ				
		プログラムA	プログラムB	プログラムC	プログラムD	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮

## 第4図

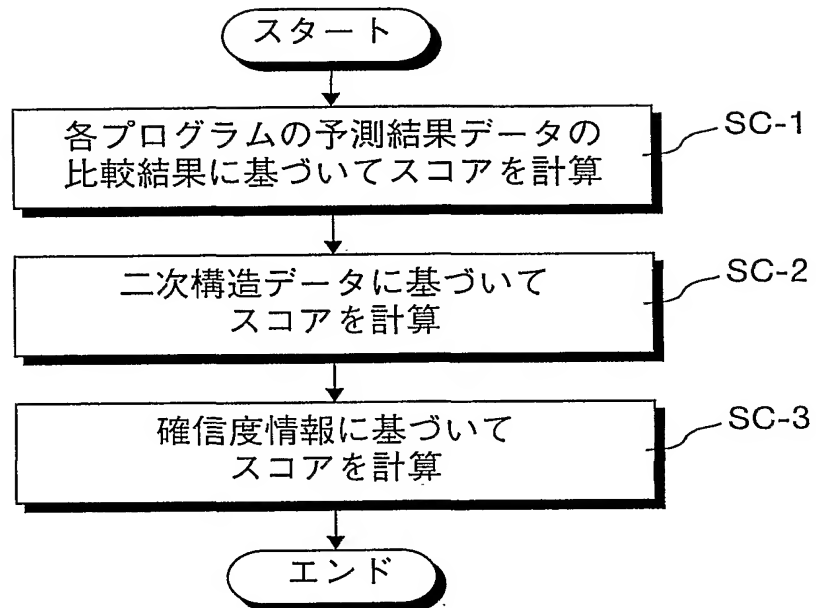


## 第5図

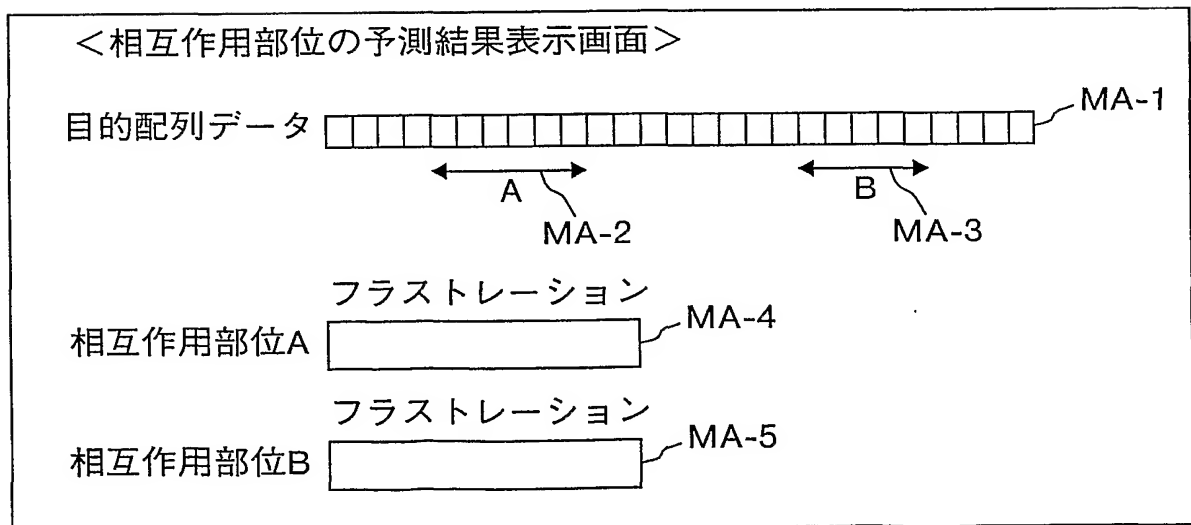


6/77

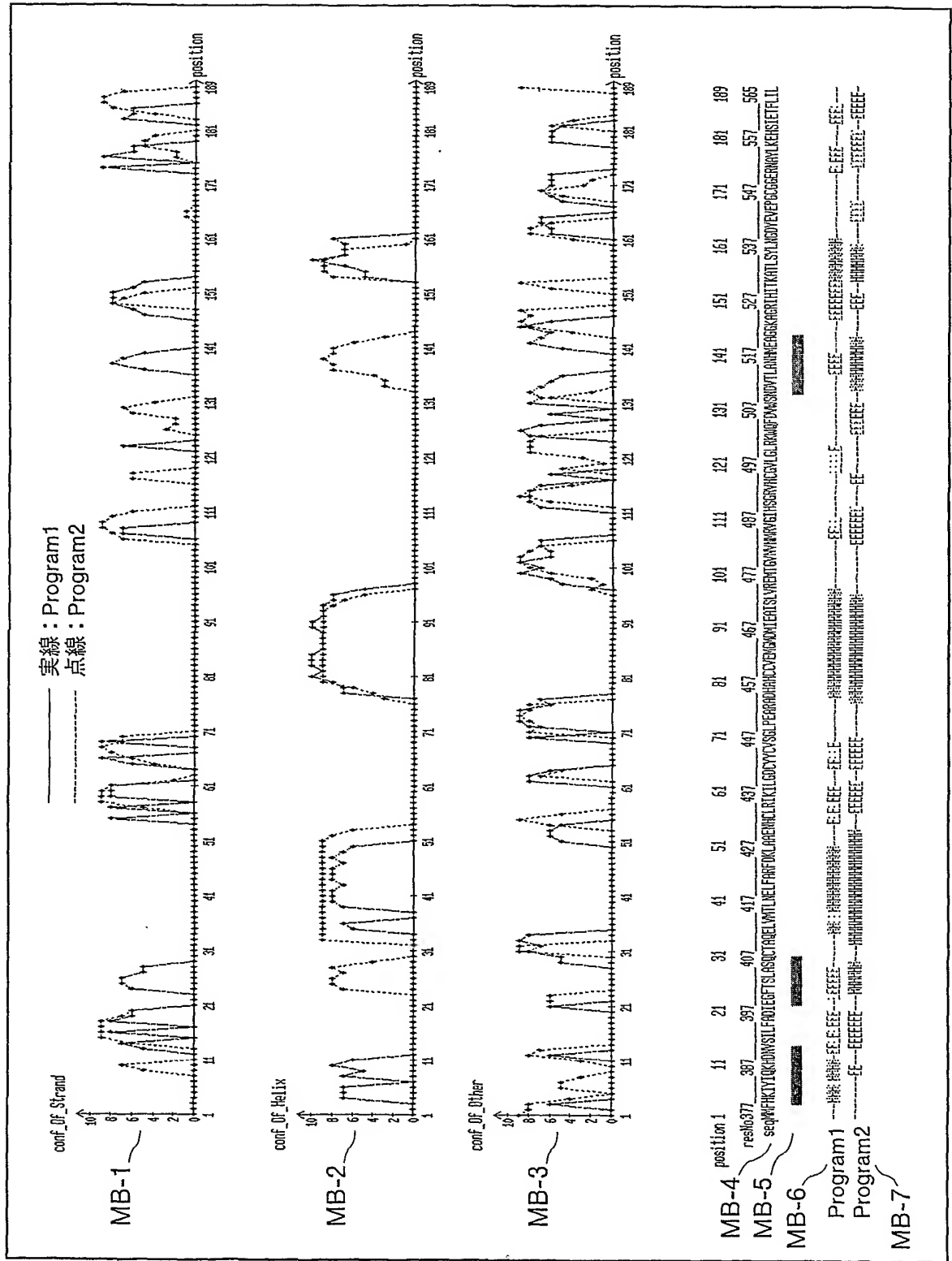
## 第6図



## 第7図

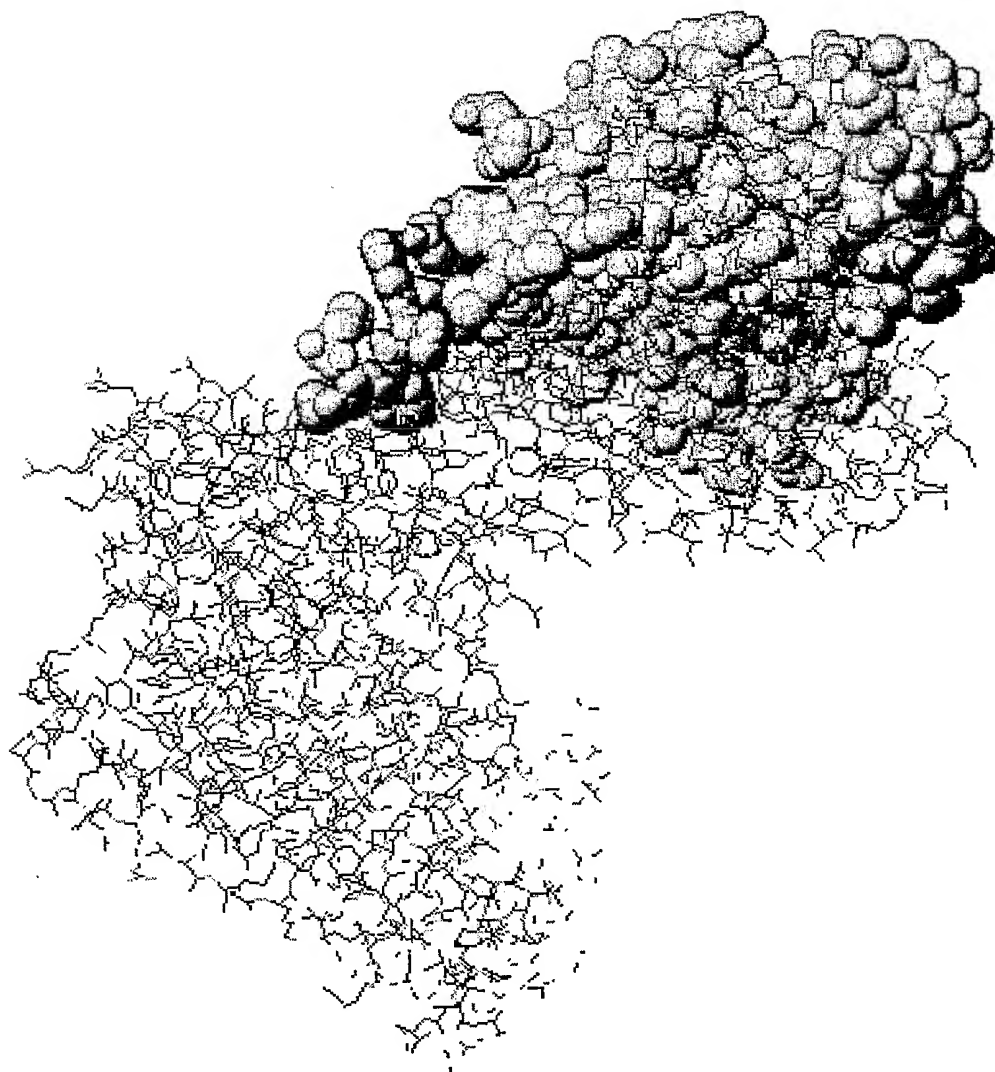


第8図

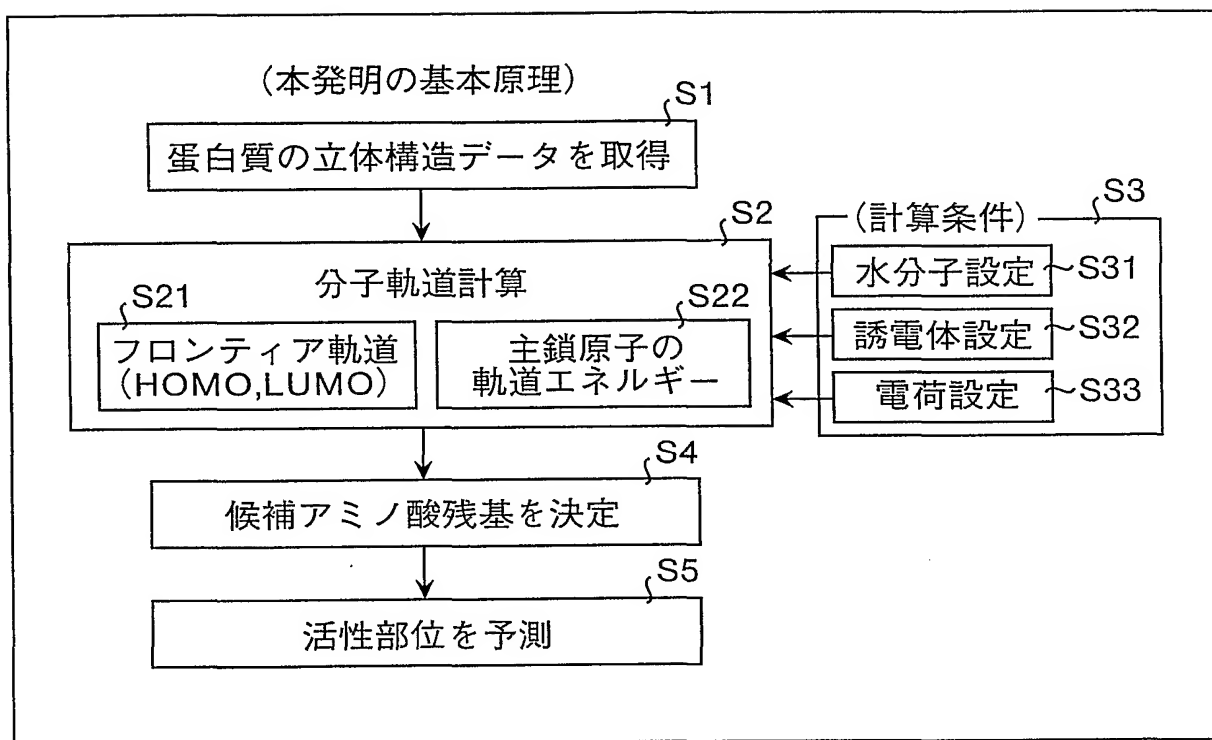




第9図

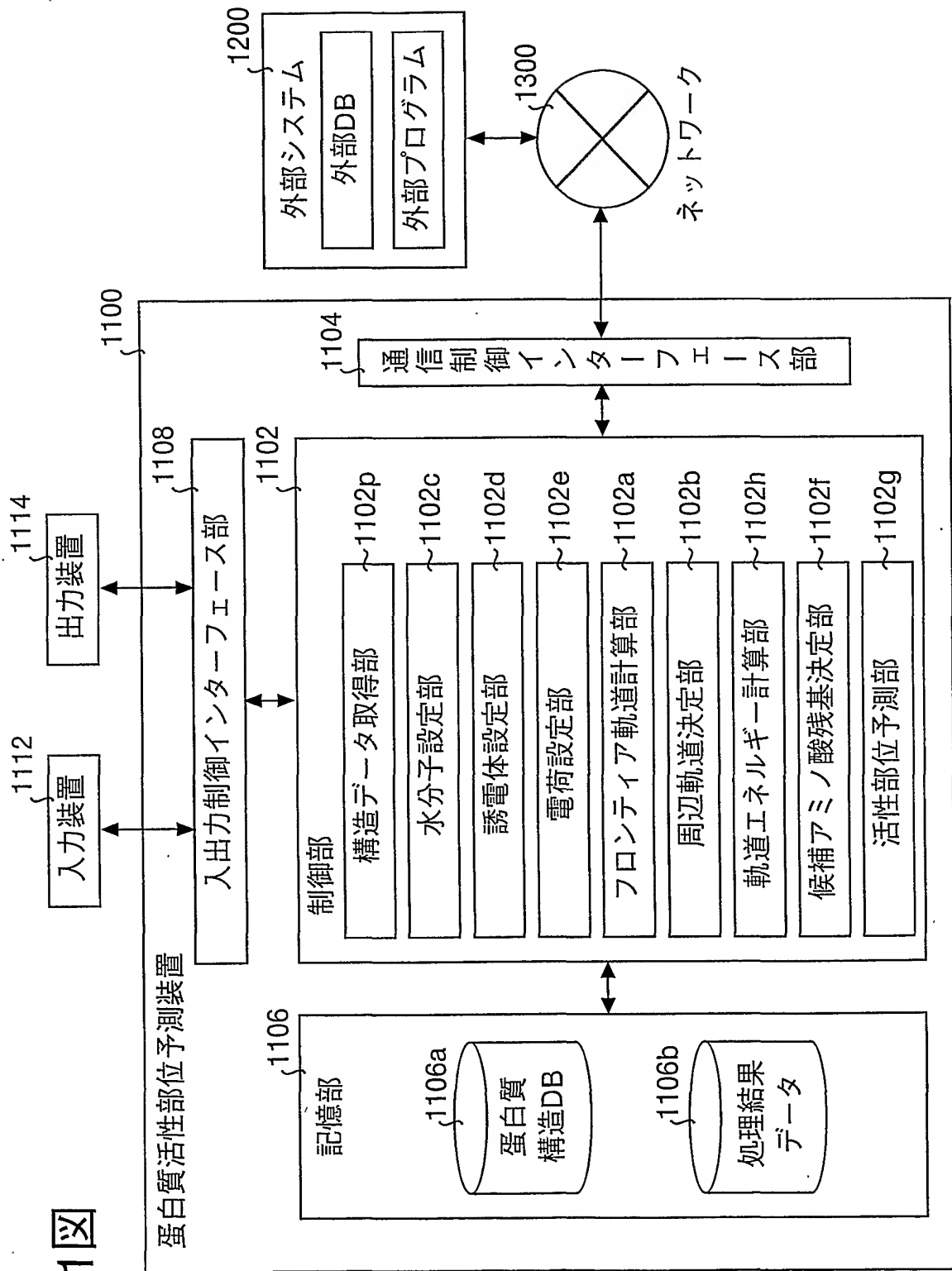


## 第10図



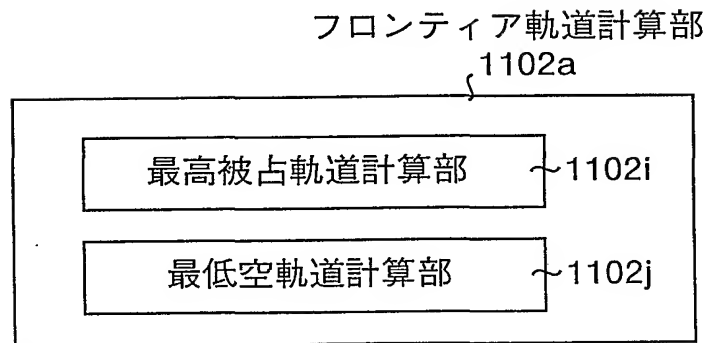
10/77

第11図

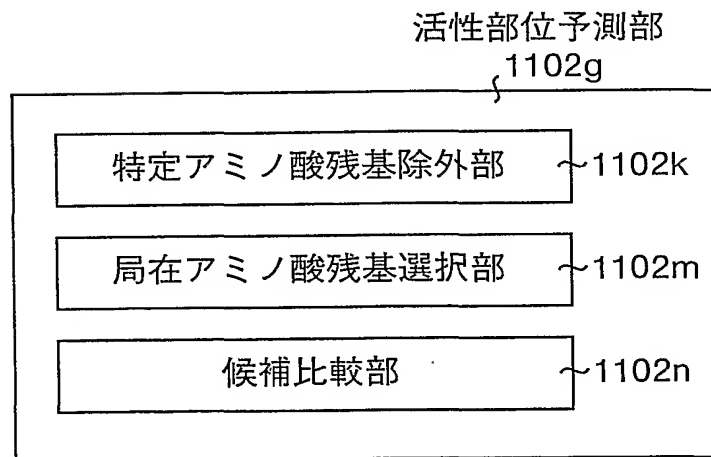


11/77

## 第12図

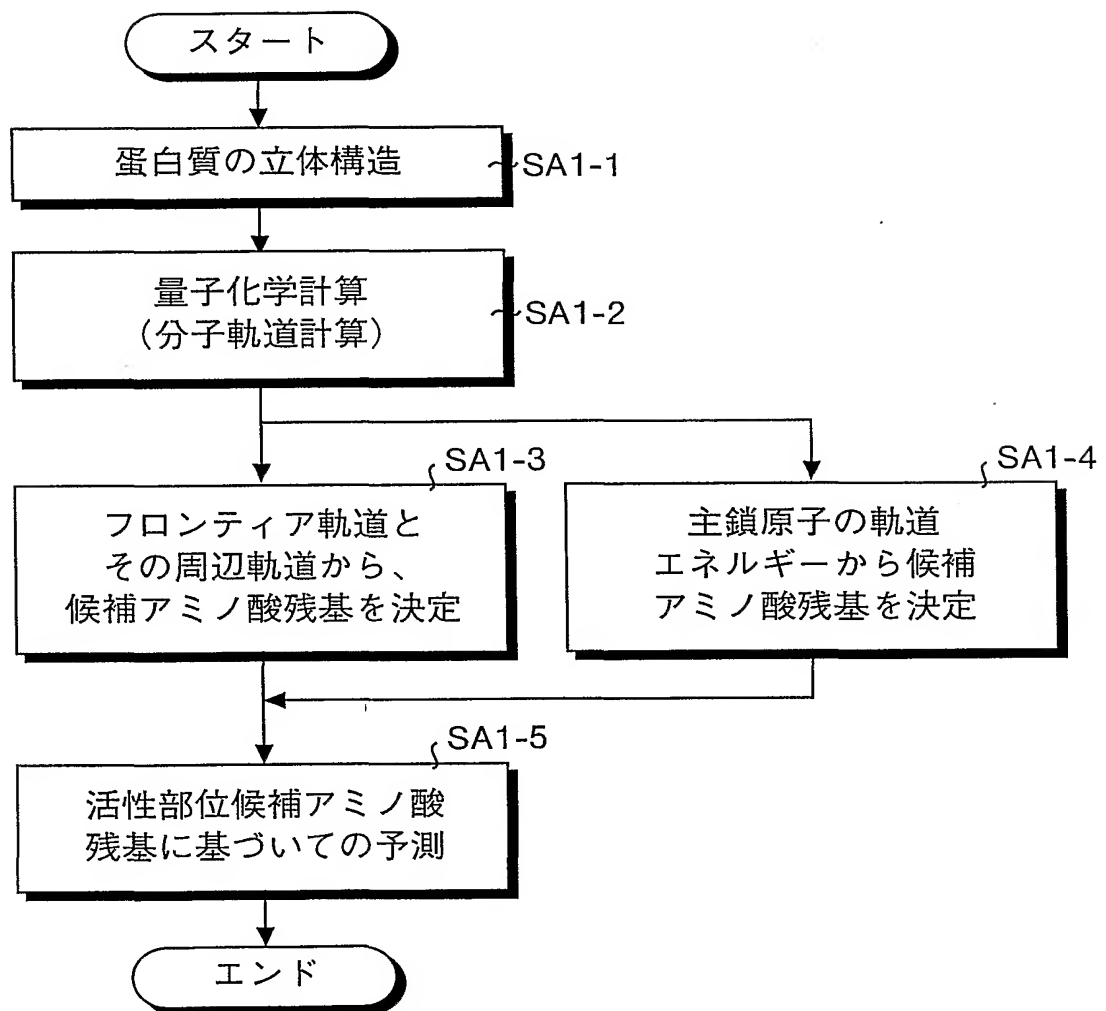


## 第13図

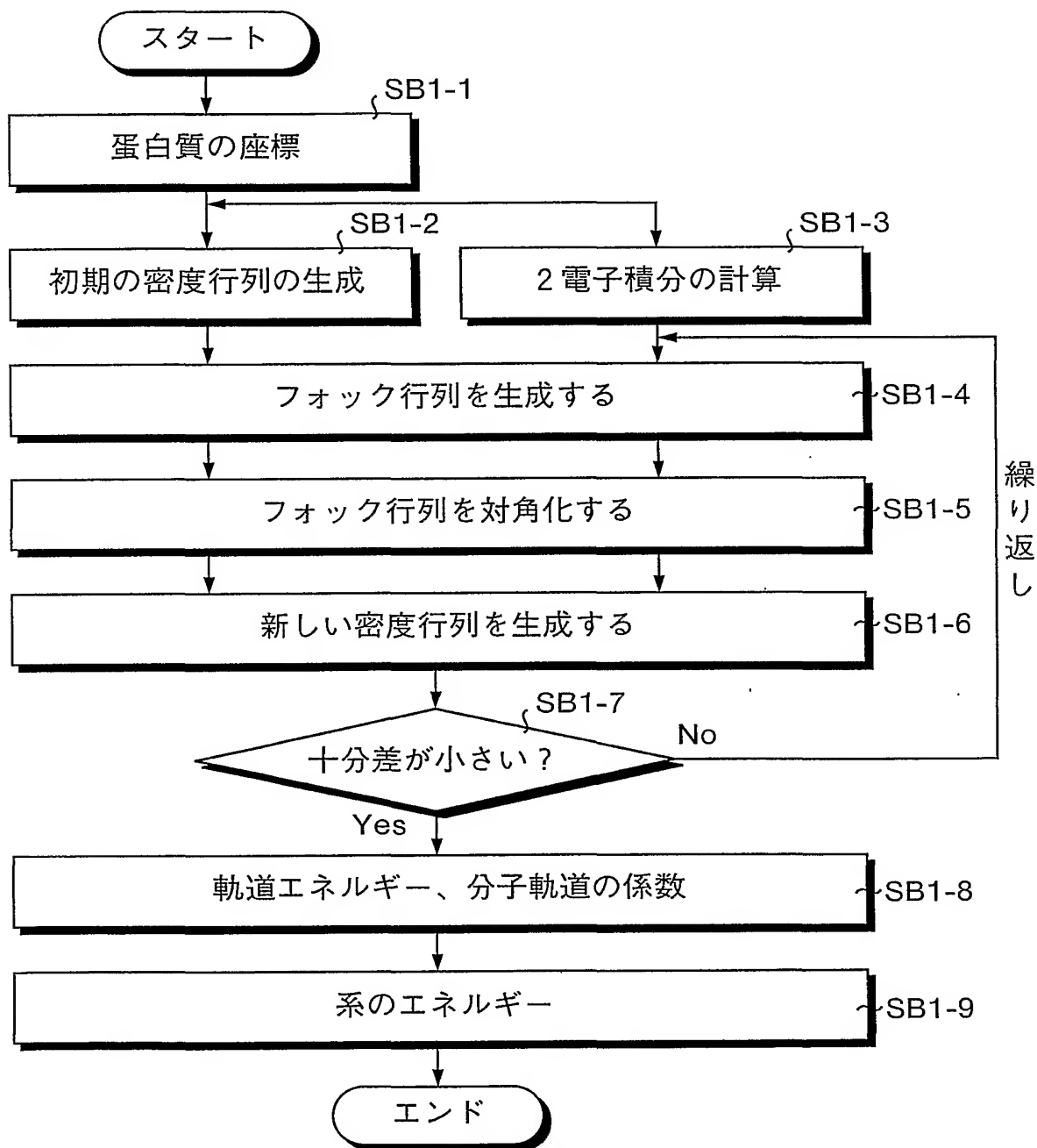


12/77

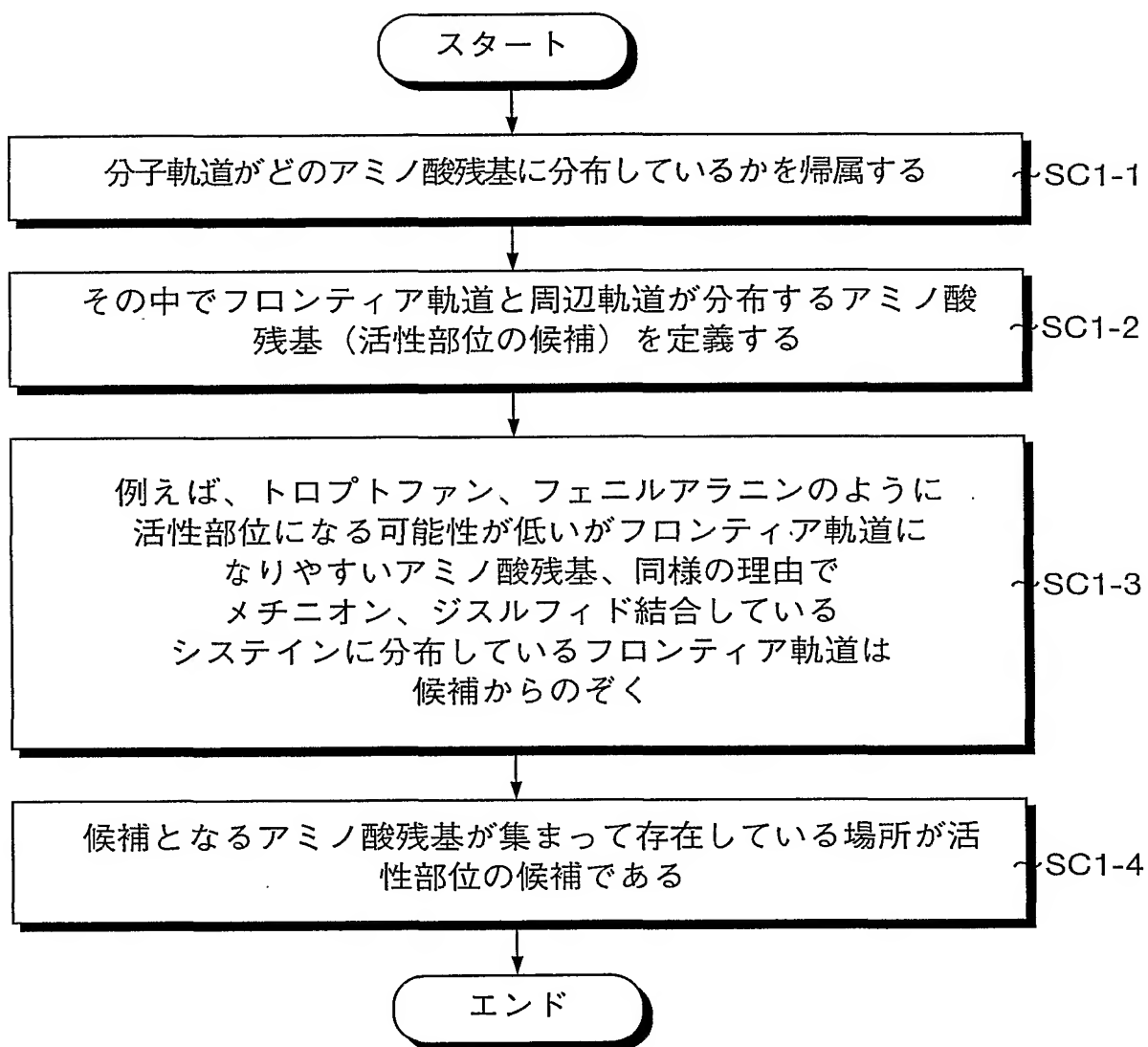
## 第14図



## 第15図

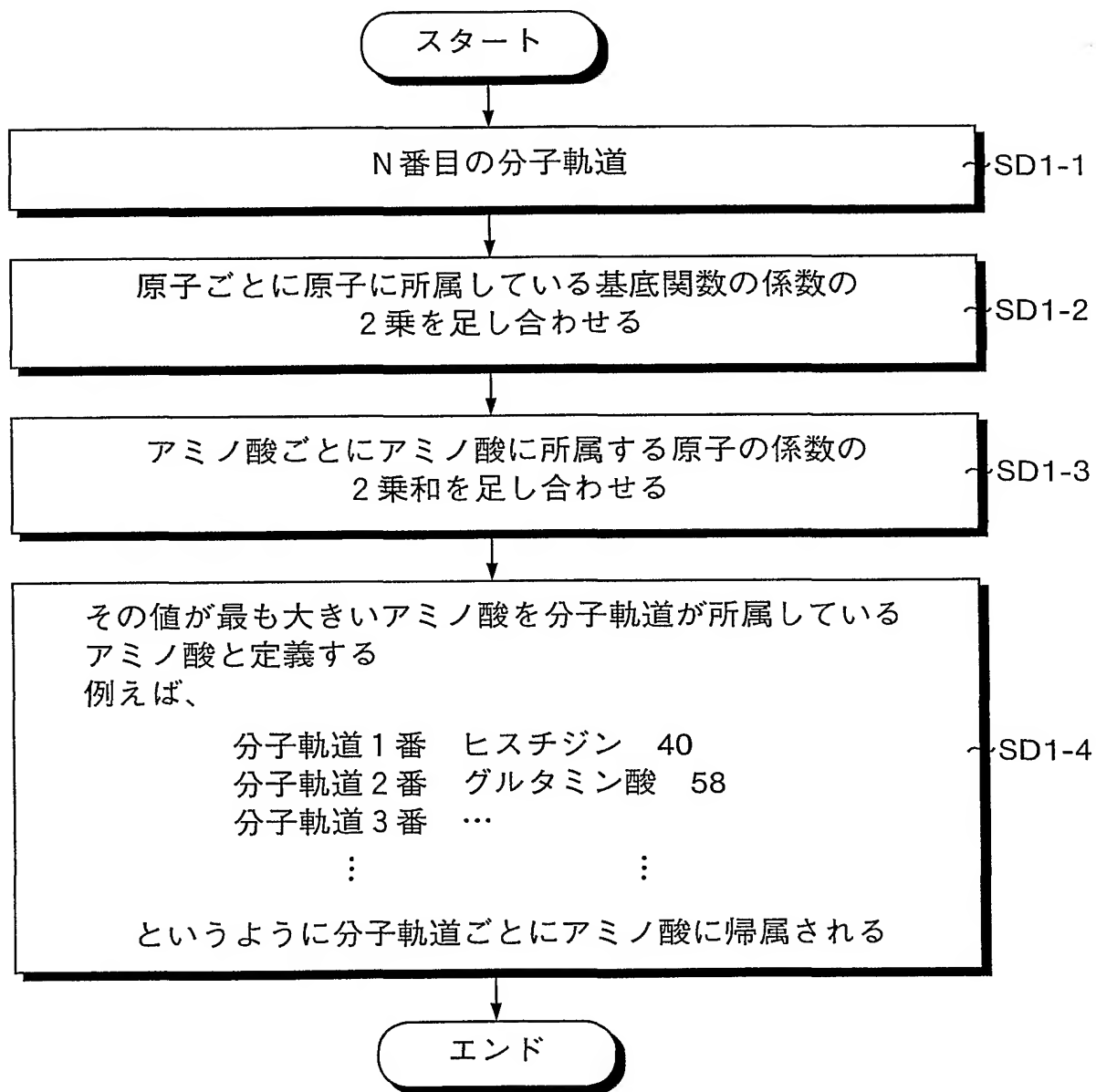


## 第16図



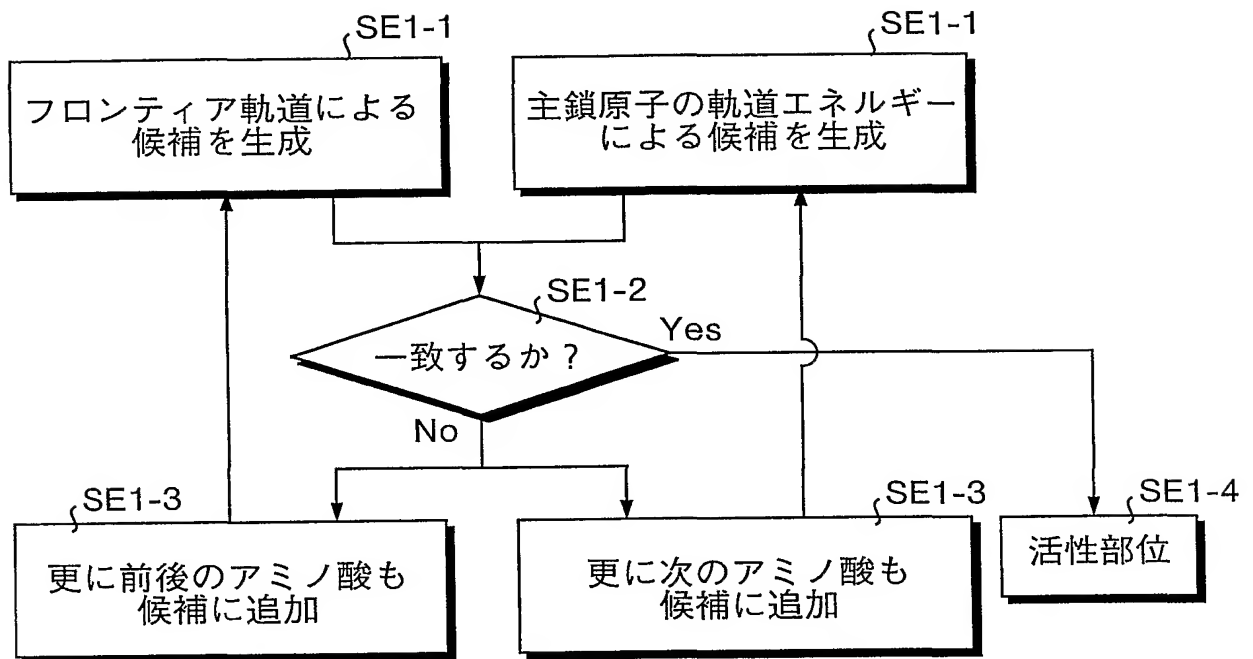
15/77

## 第17図



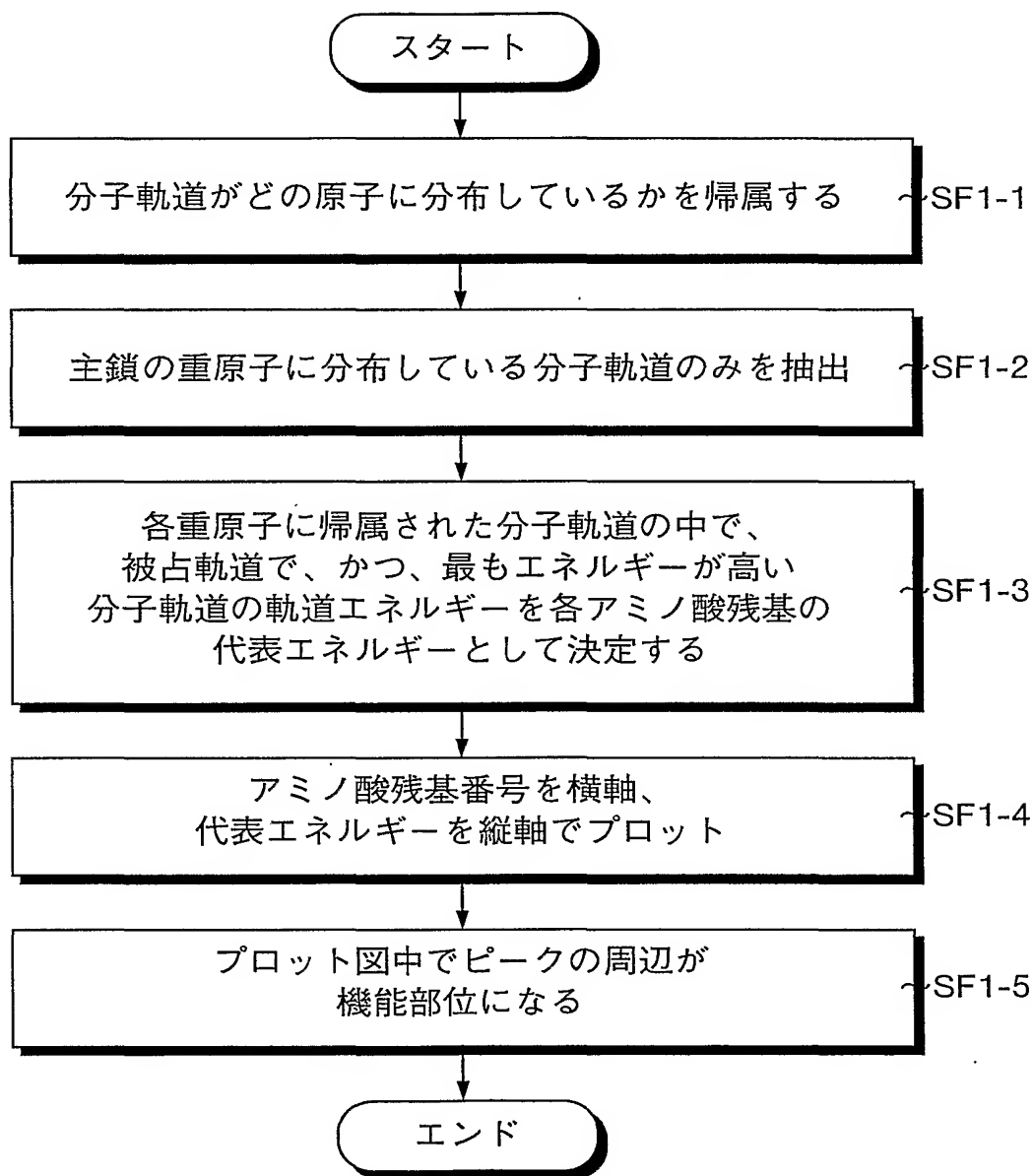


## 第18図



17/77

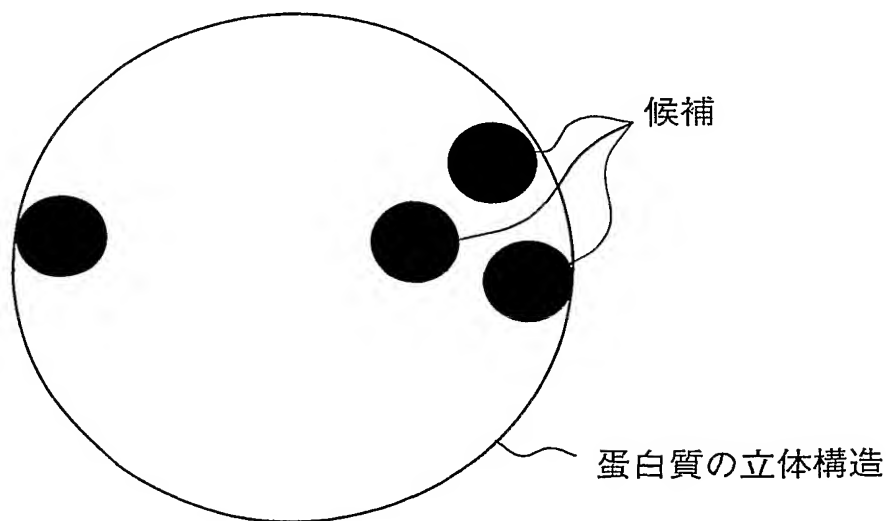
## 第19図



## 第20図

分子軌道	軌道 エネルギー	種類	分布率				
			R	E	W	T	Y
1	-10	被占	90%	1%	9%		
2	-8.5	被占	1%	4%	5%	65%	25%
3	-8	被占	2%	95%	3%		
4	-7	被占		2%	90%	8%	
5	1	空	99%		1%		
6	2	空				10%	90%
7	4	空		100%			

## 第21図



第22図

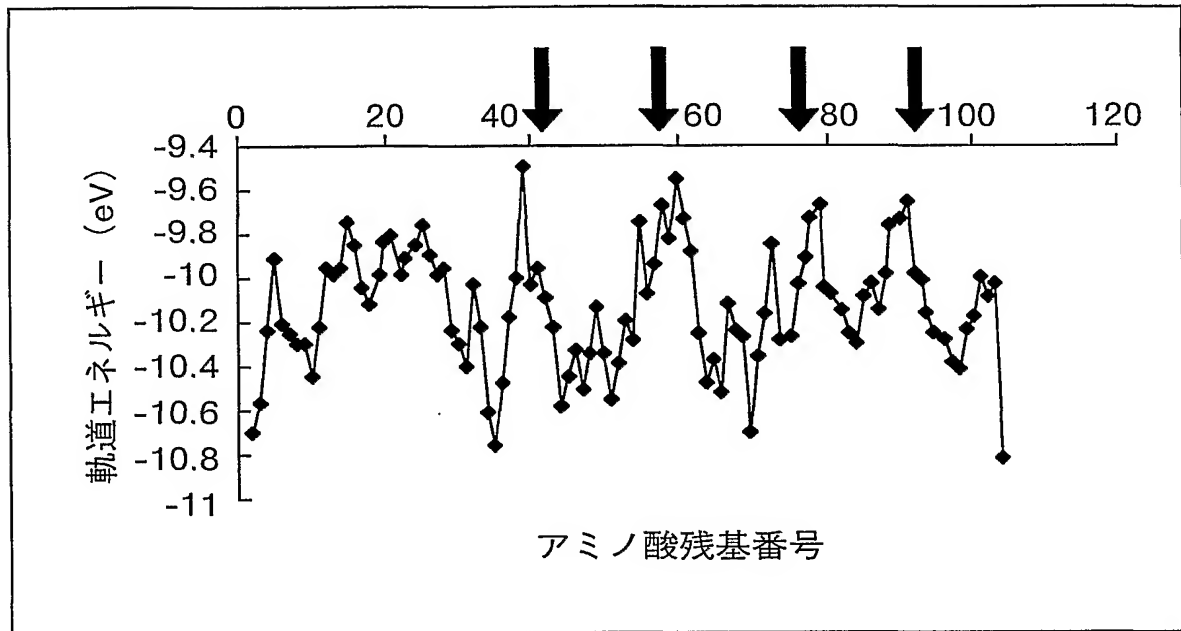
分子軌道	軌道 エネルギー	種類	R 主鎖N 原子番号1	R 側鎖C 原子番号2	R 側鎖N 原子番号3	E 主鎖N 原子番号4	その他
1	-10	被占	<u>90</u>	10			
2	-9	被占	1			<u>66</u>	33
3	-8	被占					100
4	-7	被占			5	5	90
5	-6	被占	<u>80</u>		20		
6	-5	被占				<u>60</u>	40
7	-4	被占	1	<u>99</u>			
8	-3	被占			<u>99</u>	1	
9	2	空	<u>99</u>				1
10	1	空		45		<u>50</u>	5

## 第23図

名前	軌道エネルギー	アミノ酸残基	備考	活性部位
HOMO-6	-8.734	Tyr 57	④	
HOMO-5	-8.574	Glu 58	③	○
HOMO-4	-8.543	Trp 59	×	
HOMO-3	-8.533	Tyr 11	×	
HOMO-2	-8.305	His 40	②	○
HOMO-1	-8.176	Glu 58	①	○
HOMO	-8.013	Trp 59	×	
LUMO	-1.876	Cys 2	×	
LUMO+1	-1.486	Cys 6	×	
LUMO+2	-0.328	Arg 77	①	○
LUMO+3	-0.186	His 92	②	○
LUMO+4	-0.186	Tyr 68	③	
LUMO+5	-0.186	His 27	④	

21/77

第24図



第25図

候補	アミノ酸残基	軌道エネルギー
最も高い	39	-9.508
2番目に高い	60	-9.563
3番目に高い	91	-9.665
4番目に高い	58	-9.673
5番目に高い	79	-9.681

## 第26図

フロンティア 軌道による候補	主鎖原子の 軌道エネルギーに よる候補	共通	活性部位
27	77-81	40	40
40	56-60	57	58
57	89-93	58	77
58	58-62	77	92
77	37-41	92	
68			
92			

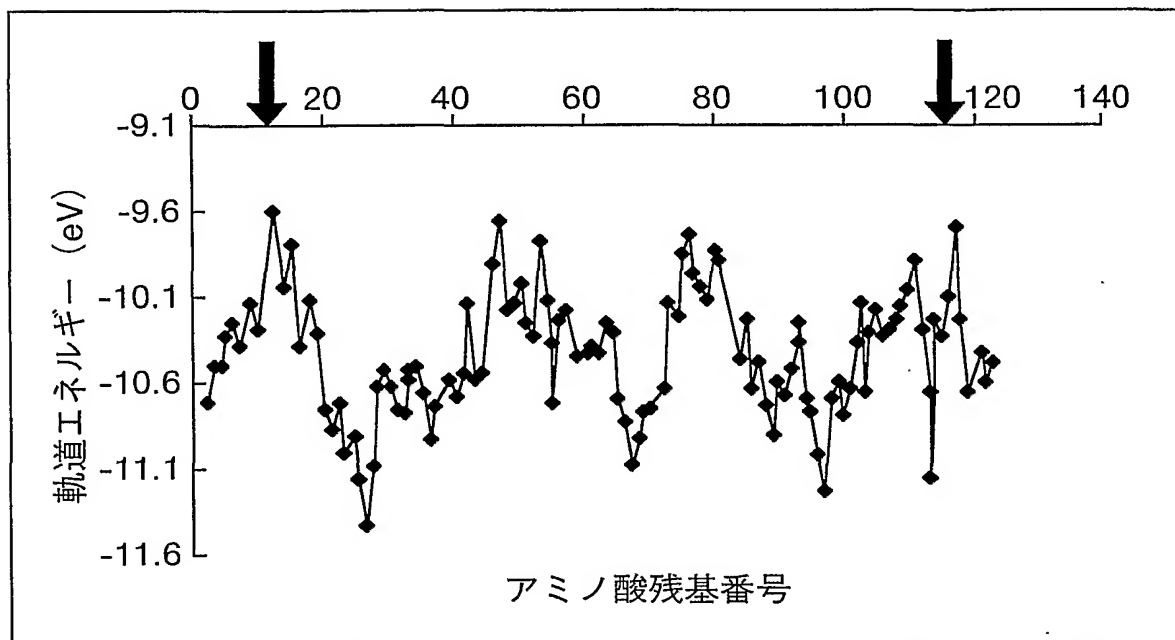
## 第27図

名前	軌道エネルギー	アミノ酸残基	備考	活性部位
HOMO-5	-8.699	Tyr 73	④	
HOMO-4	-8.604	His 48	③	
HOMO-3	-8.545	Met 13	×	
HOMO-2	-8.424	His 12	②	○
HOMO-1	-8.420	Asp 14	①	
HOMO	-8.353	Met 79	×	
LUMO	-2.457	Cys 84	×	
LUMO+1	-2.233	Cys 72	×	
LUMO+2	-2.122	Cys 40	×	
LUMO+3	-1.575	Cys 58	×	
LUMO+4	-0.489	Tyr 97	①	
LUMO+5	-0.428	Cys 26	×	
LUMO+6	-0.274	Tyr 97	①	
LUMO+7	-0.124	Cys 26	×	
LUMO+8	-0.085	Tyr 25	②	
LUMO+9	-0.075	Arg 85	③	
LUMO+10	-0.047	Cys 95	×	
LUMO+11	-0.024	His 119	④	○



24/77

第28図



第29図

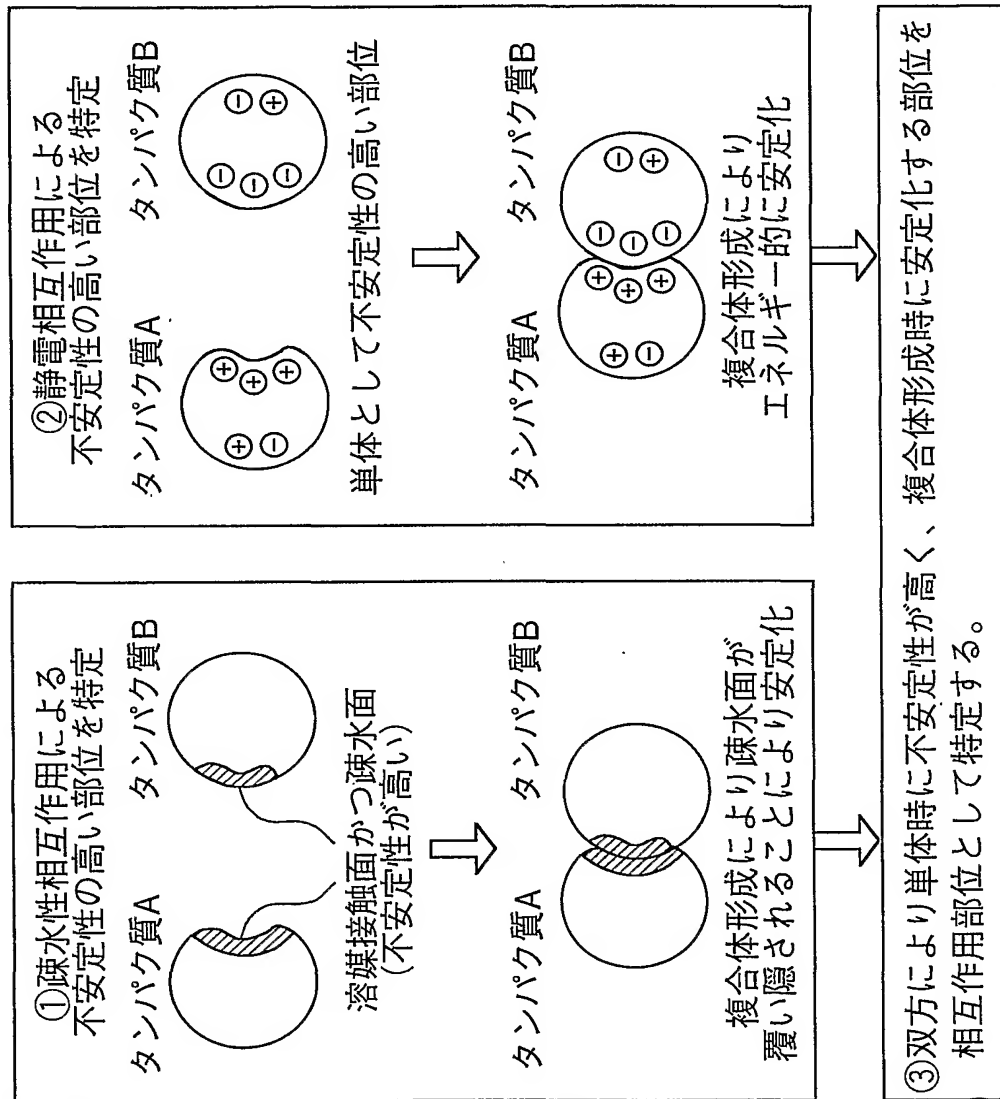
候補	アミノ酸残基	軌道エネルギー
1 番目に高い	12	-9.533
2 番目に高い	47	-9.59
3 番目に高い	117	-9.616
4 番目に高い	76	-9.669
5 番目に高い	53	-9.715

## 第30図

フロンティア 軌道による候補	主鎖原子の 軌道エネルギー による候補	共通	活性部位
12	10-14	12	12
14	45-49	14	119
25	115-119	119	
48	74-78		
73	51-55		
85			
97			
119			

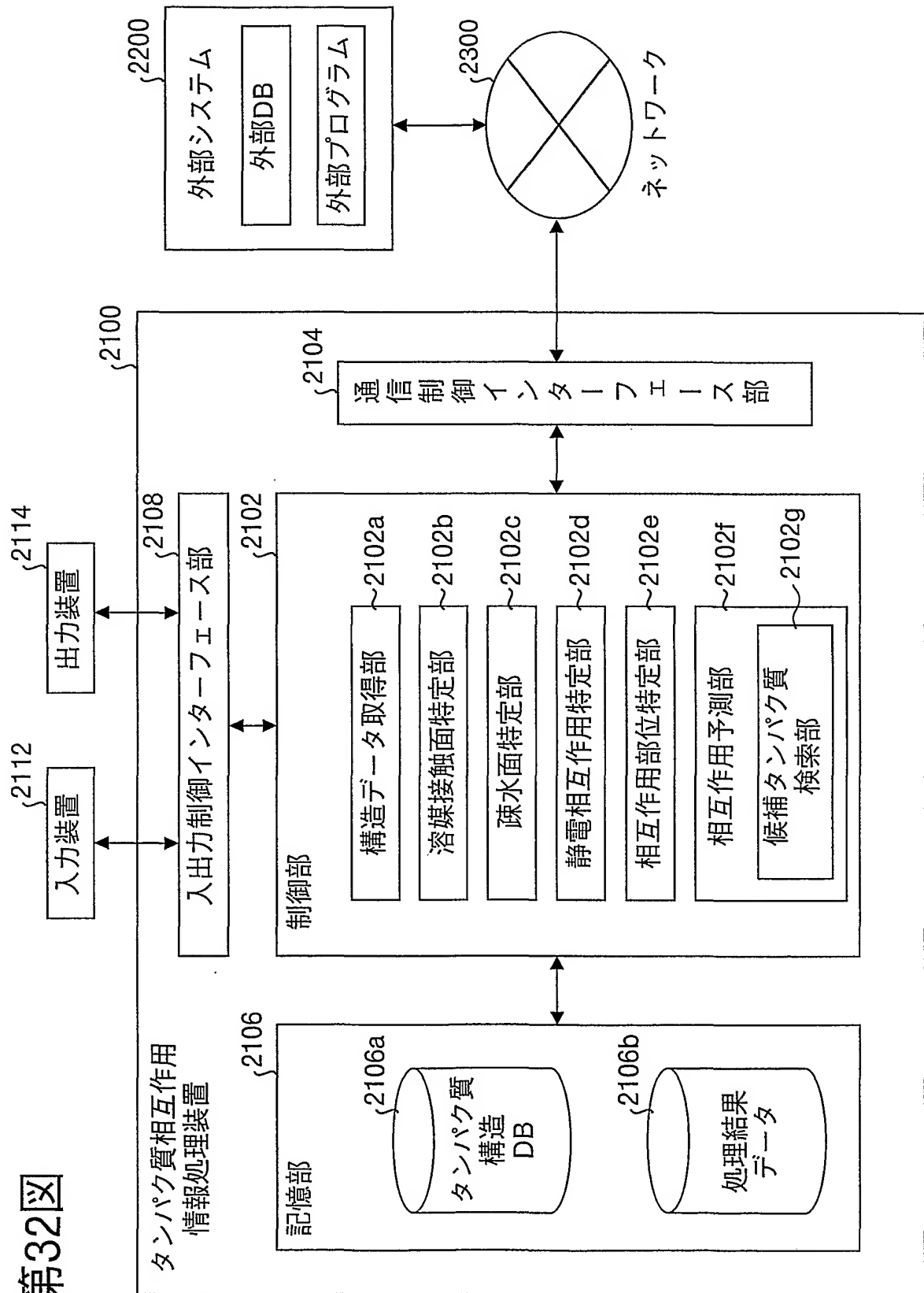
## 第31図

## (本発明の基本原理)



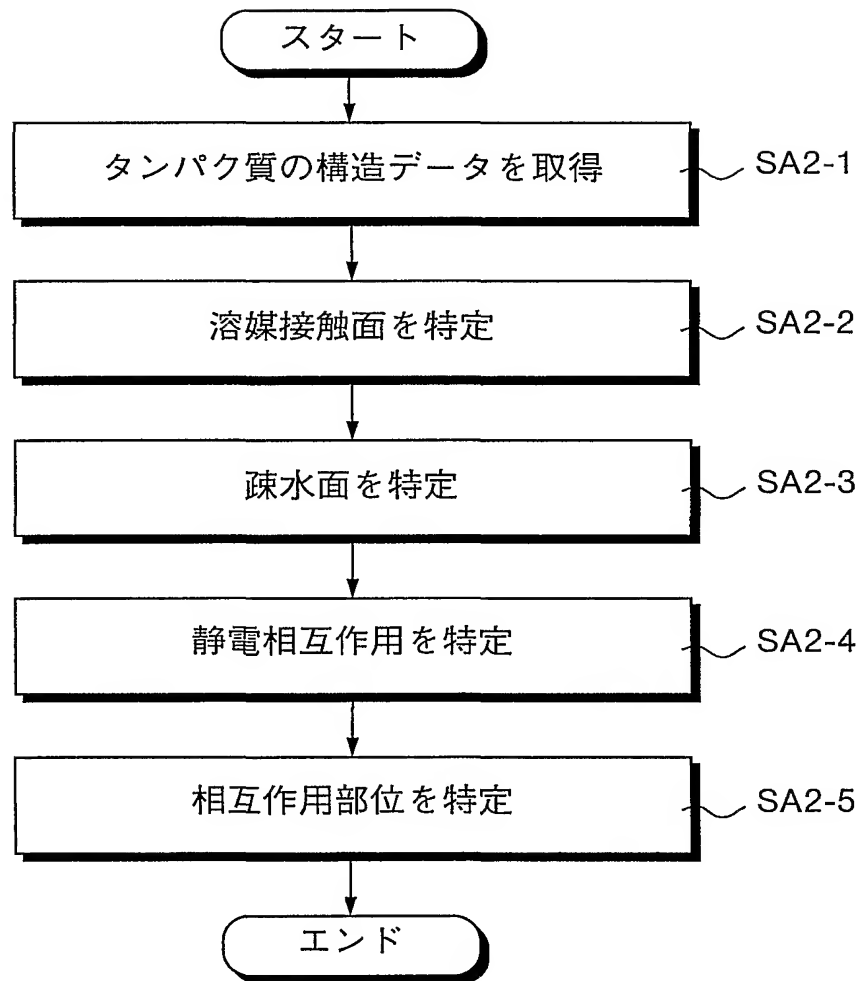
27/77

第32図



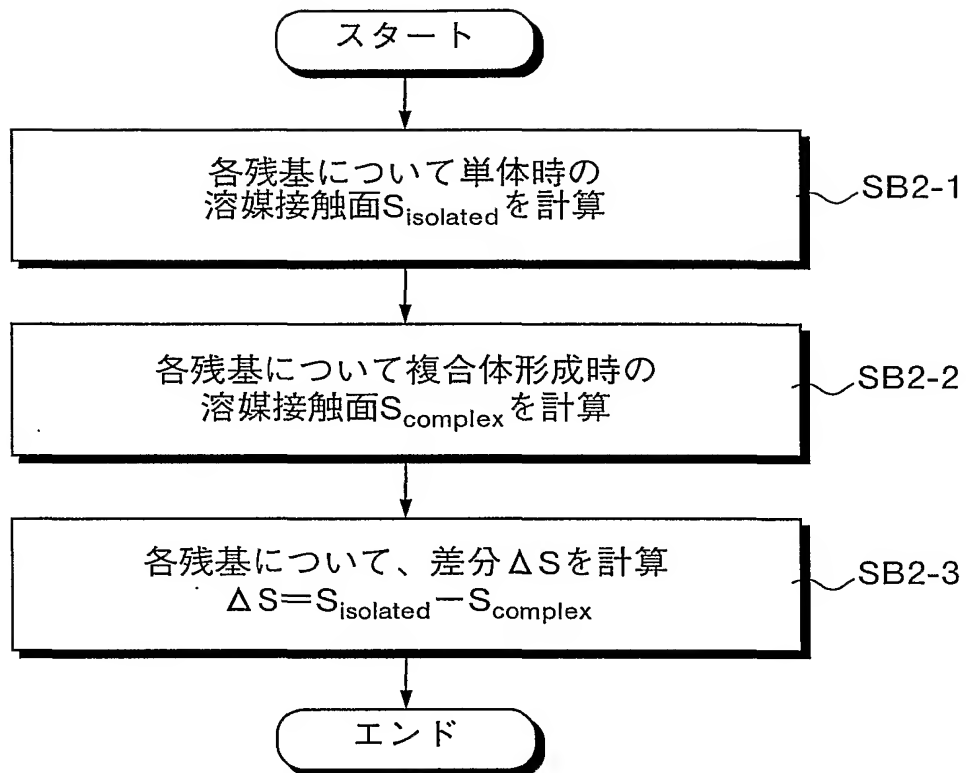
28/77

## 第33図



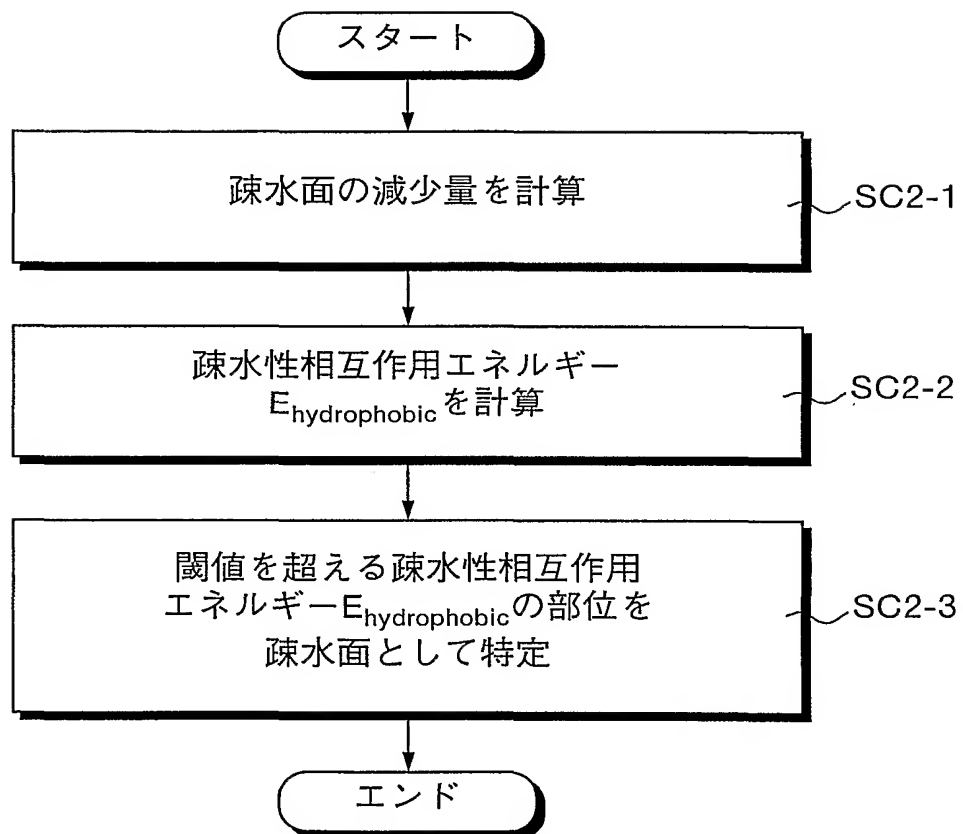
29/77

## 第34図



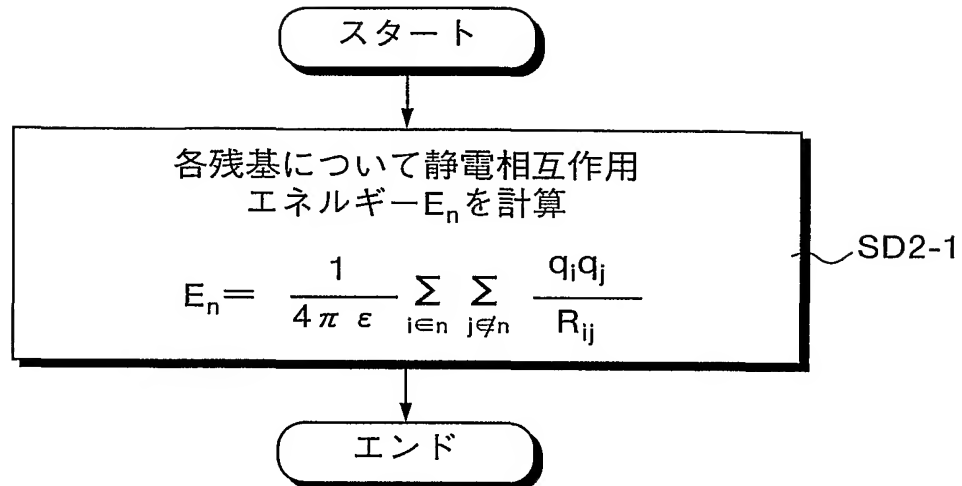
30/77

## 第35図

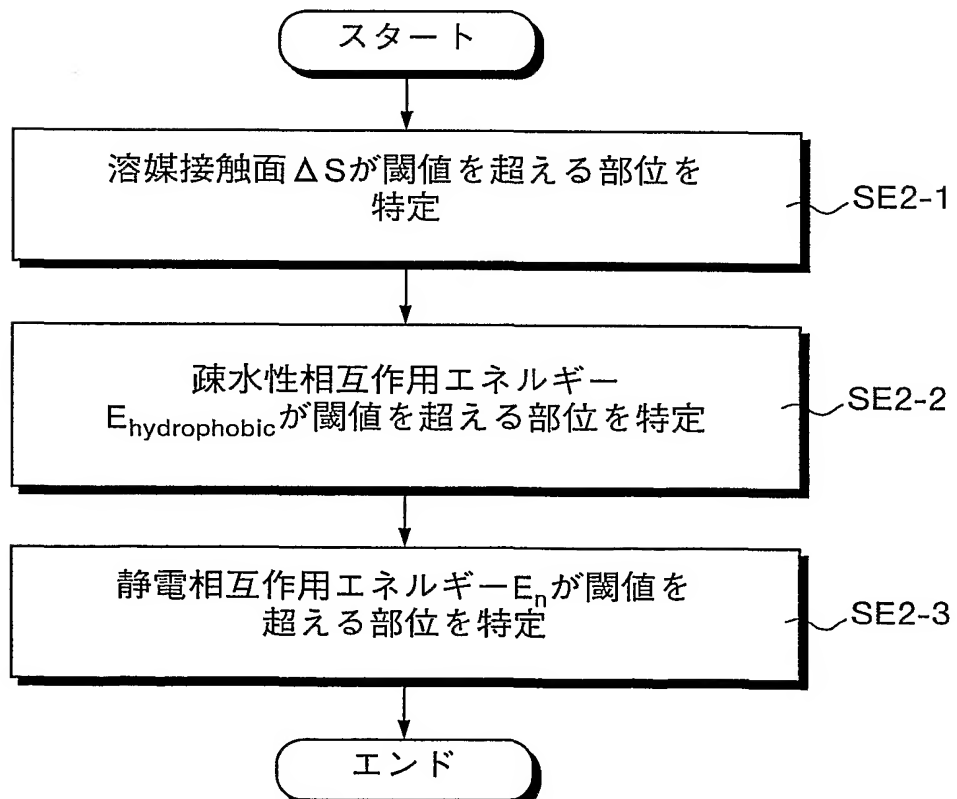


31/77

## 第36図



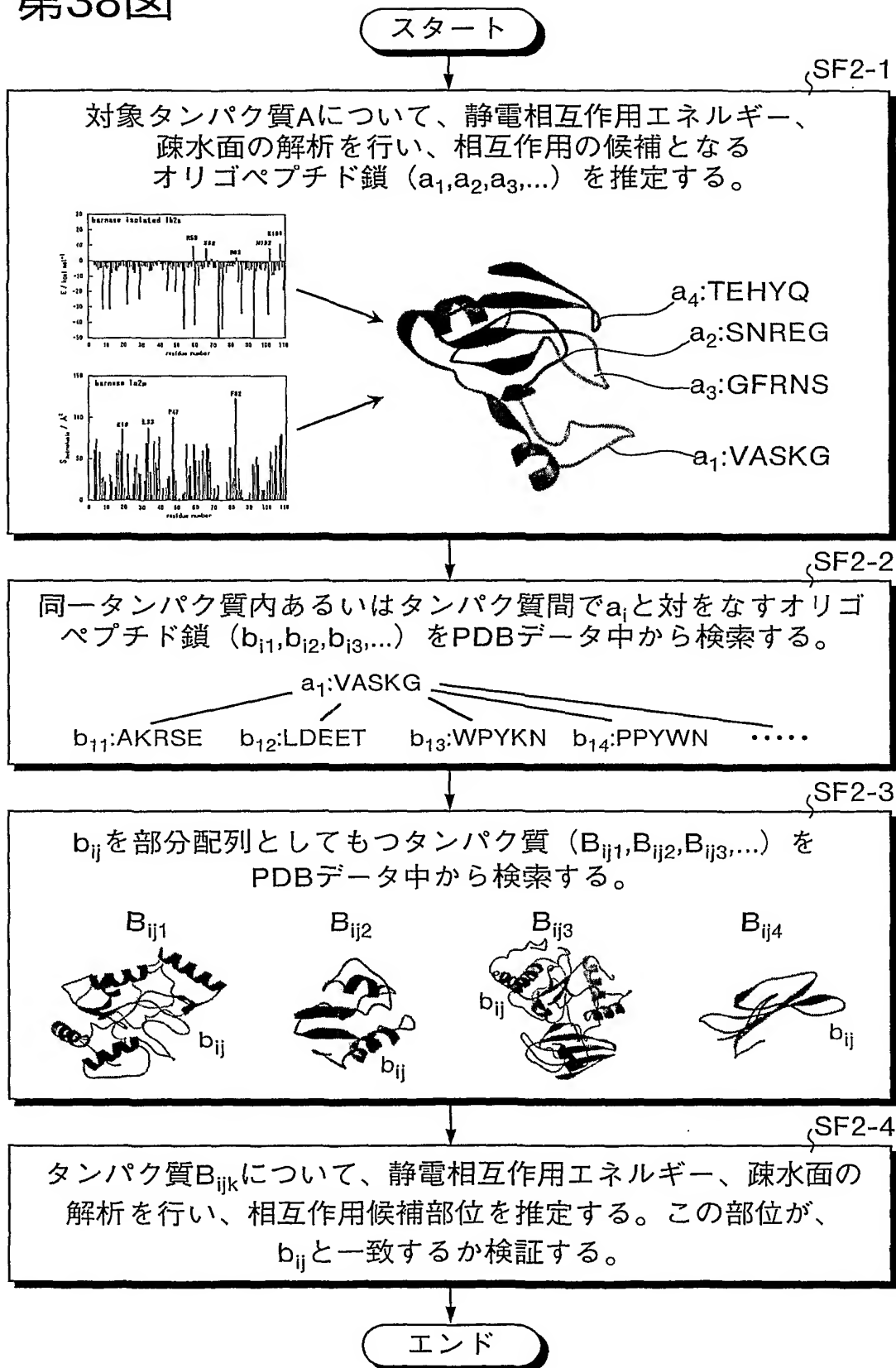
## 第37図





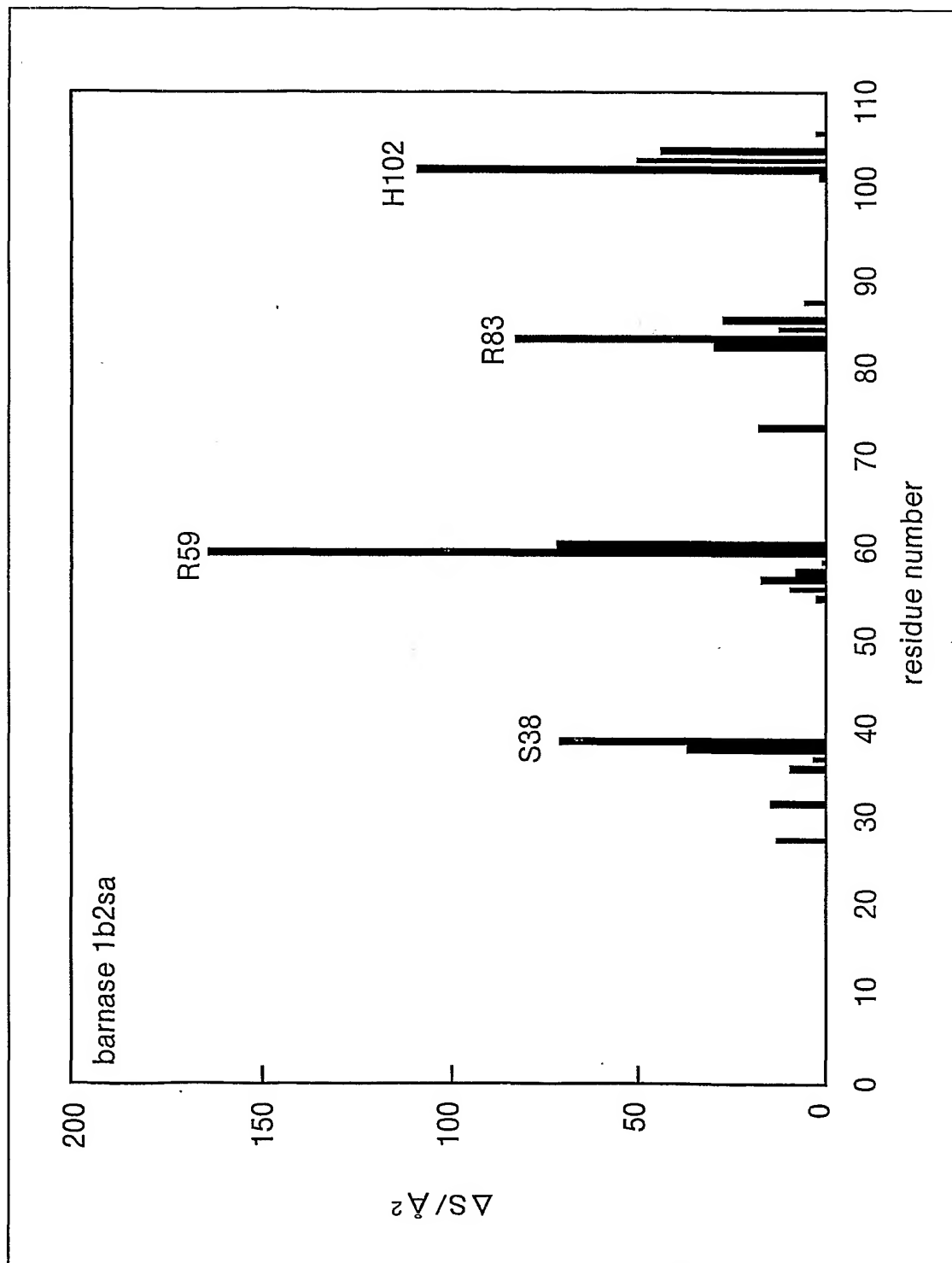
32/77

## 第38図



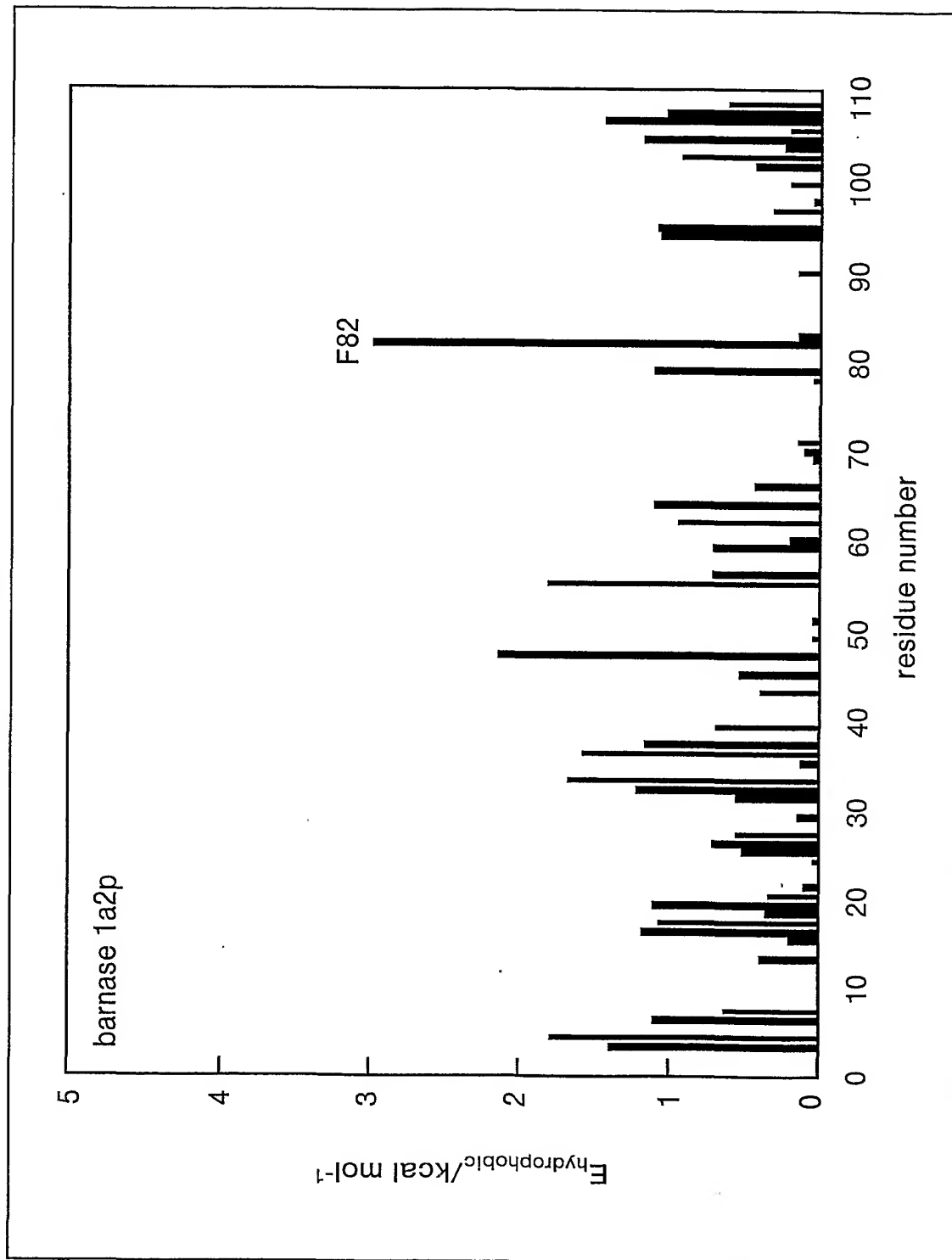
33/77

第39図



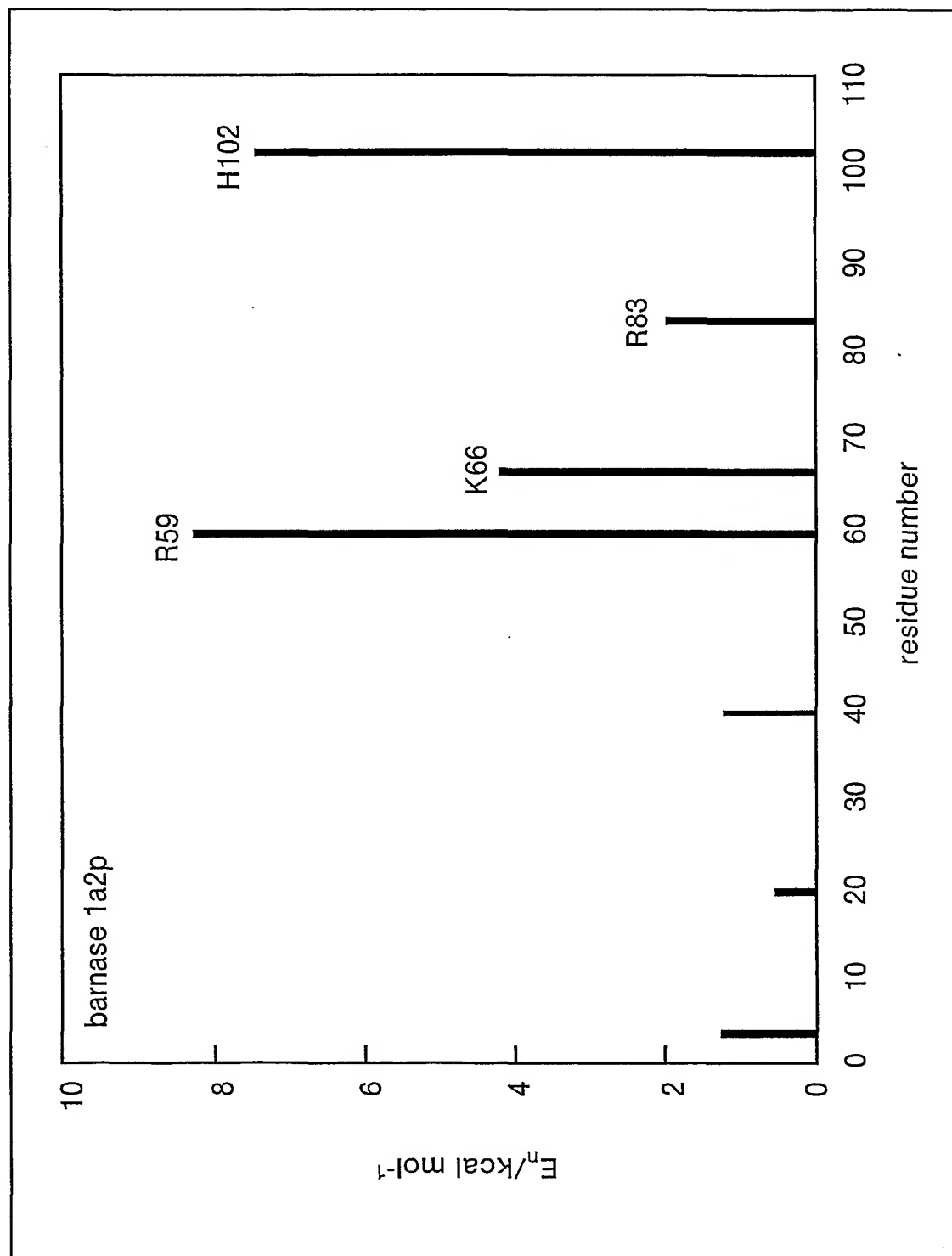
34/77

第40図

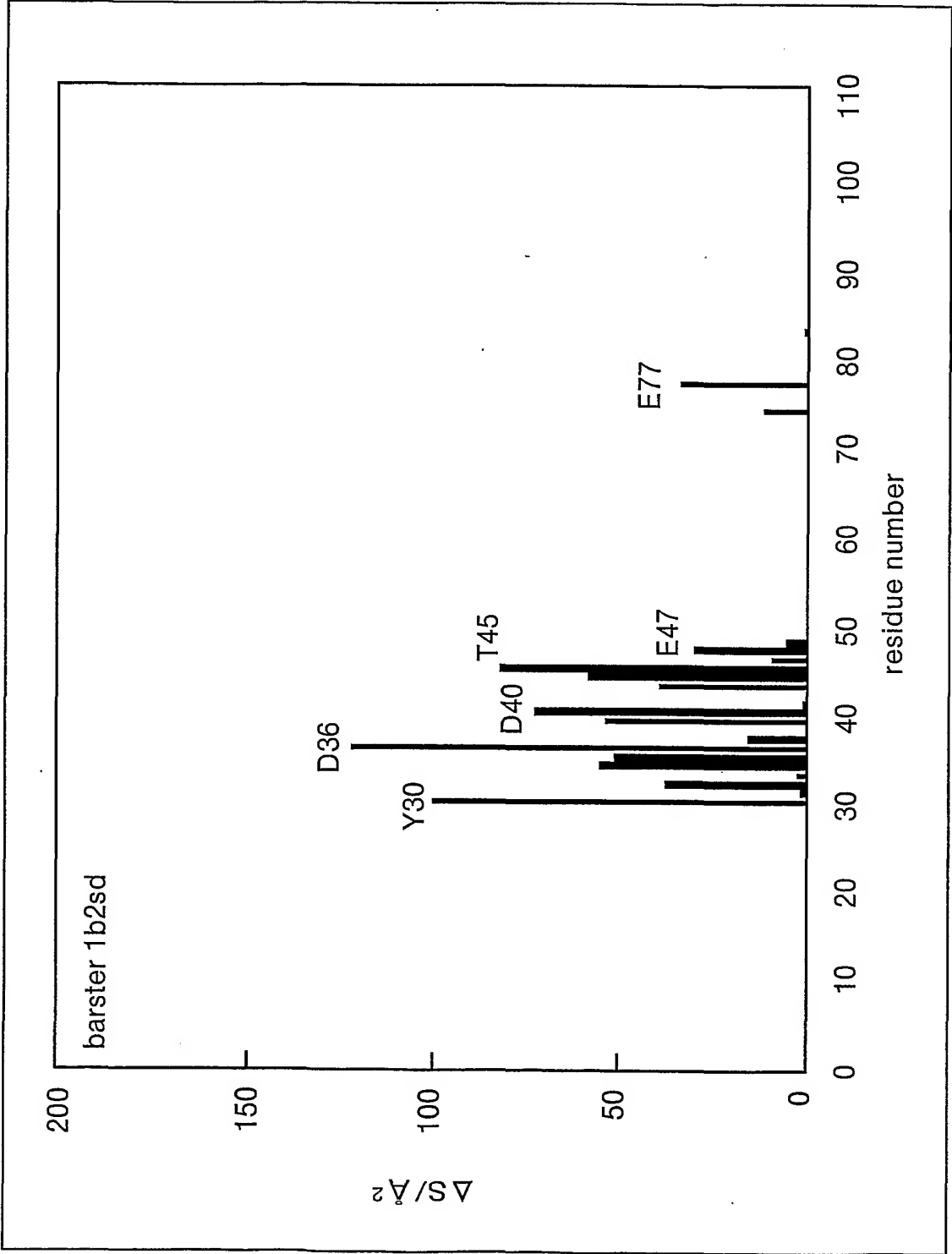


35/77

第41図

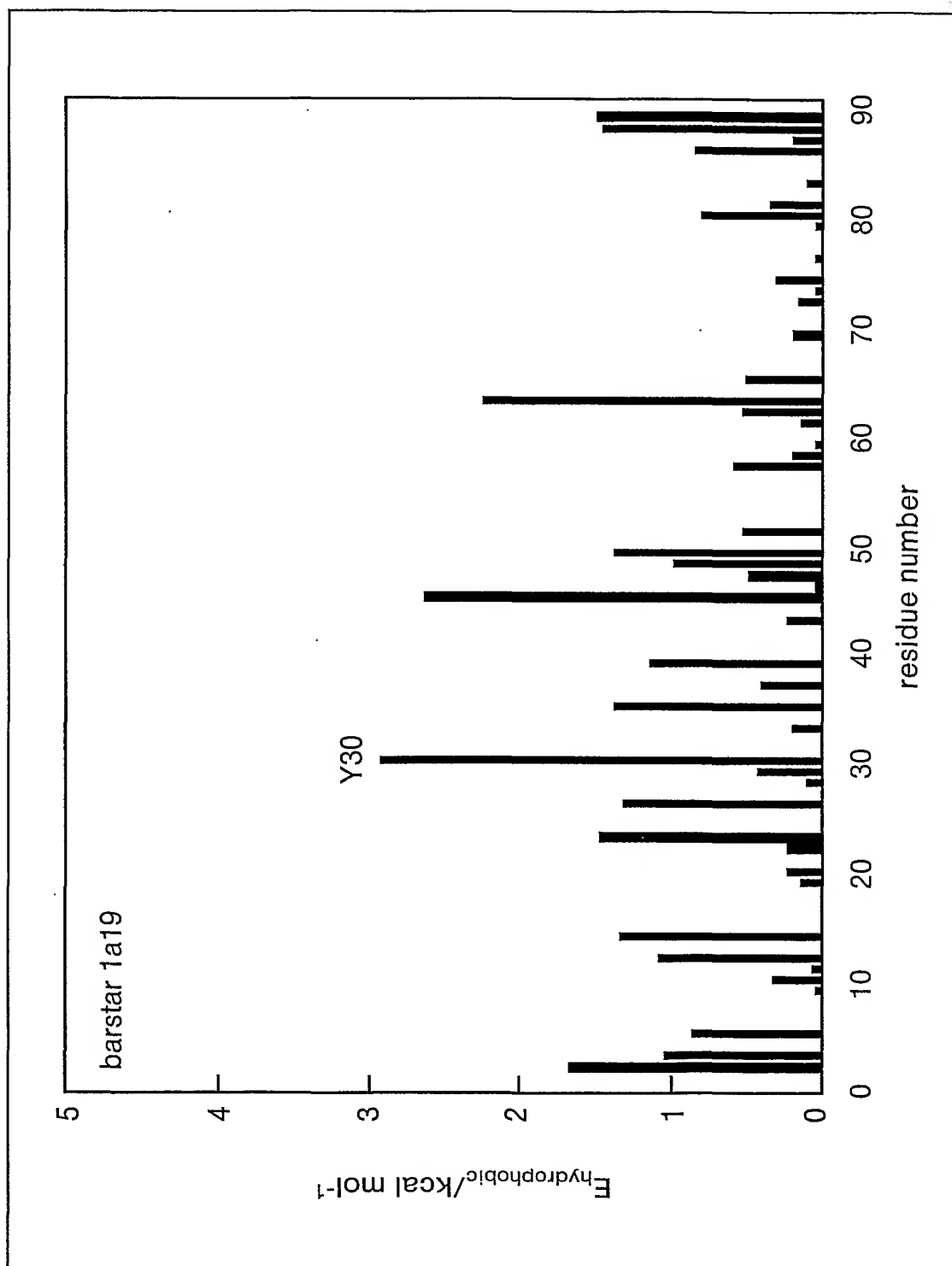


第42図



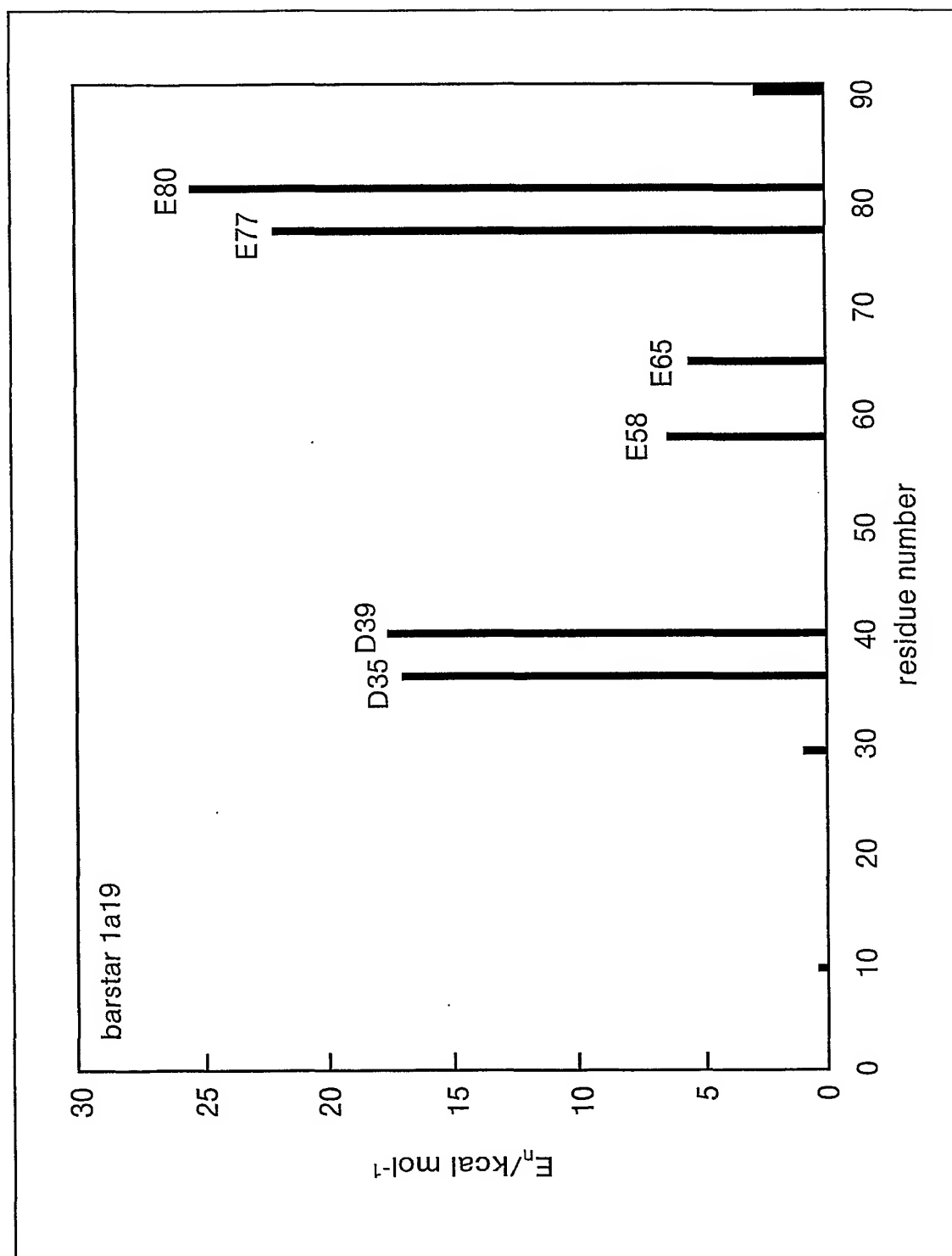
37/77

第43図



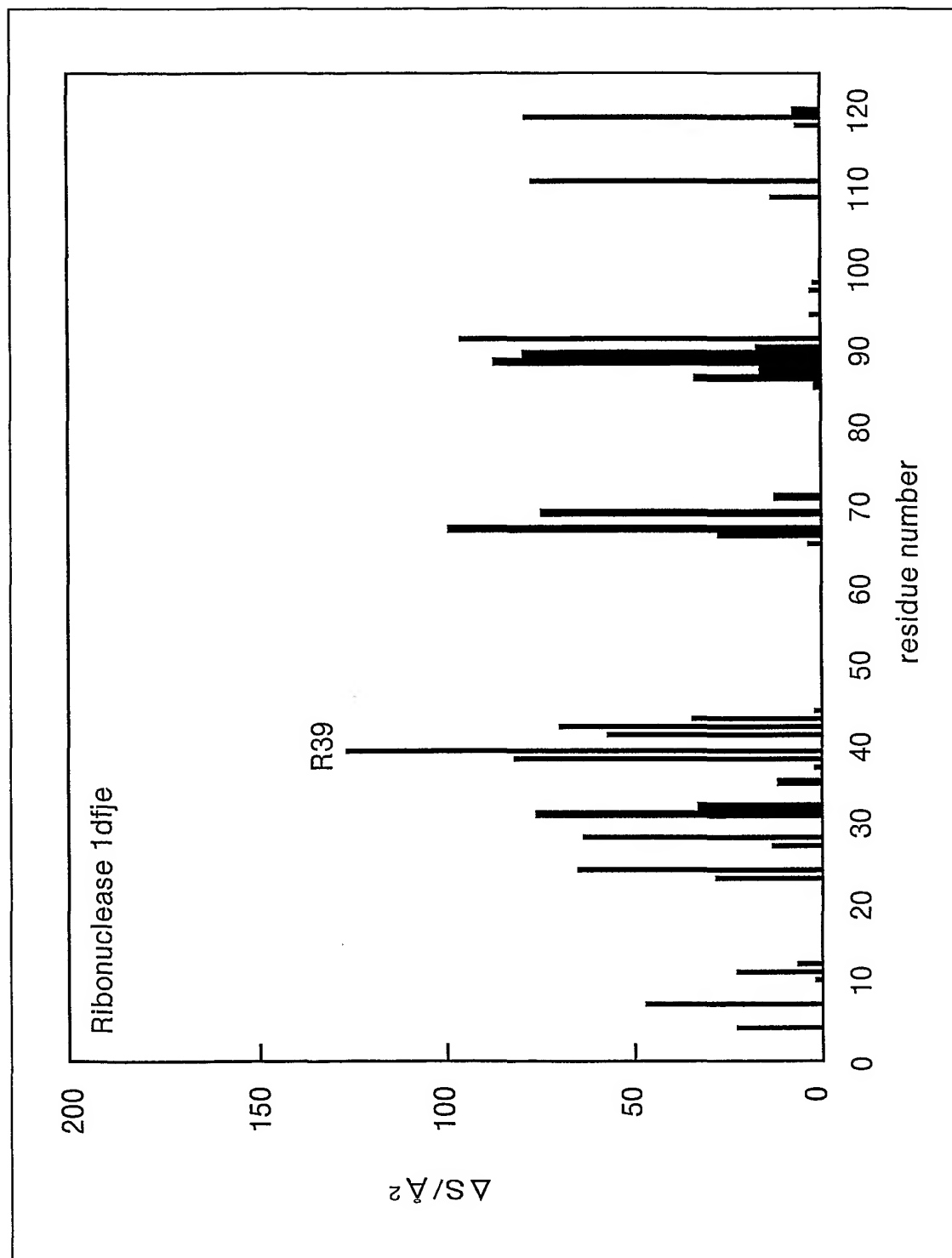
38/77

第44図



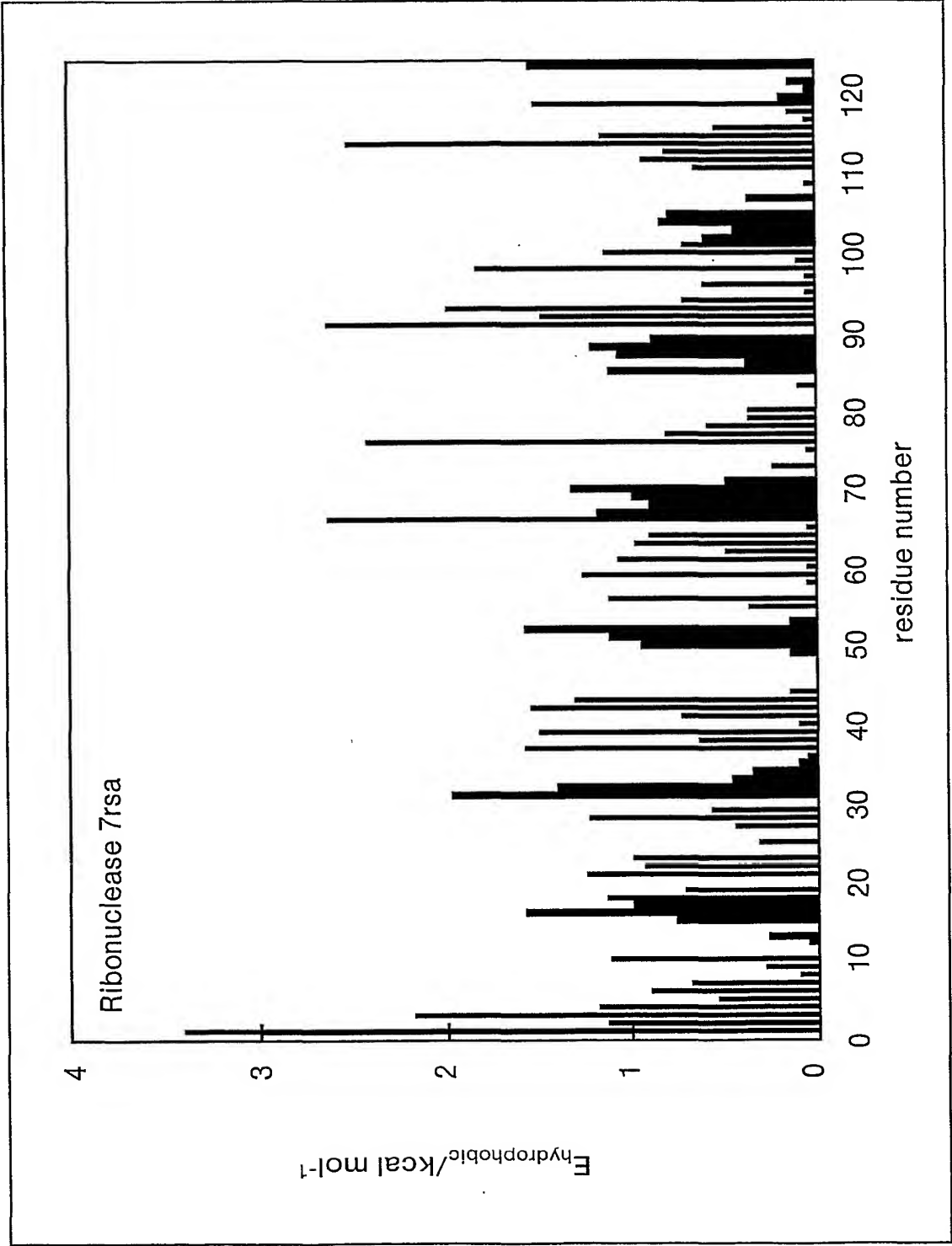
39/77

第45図

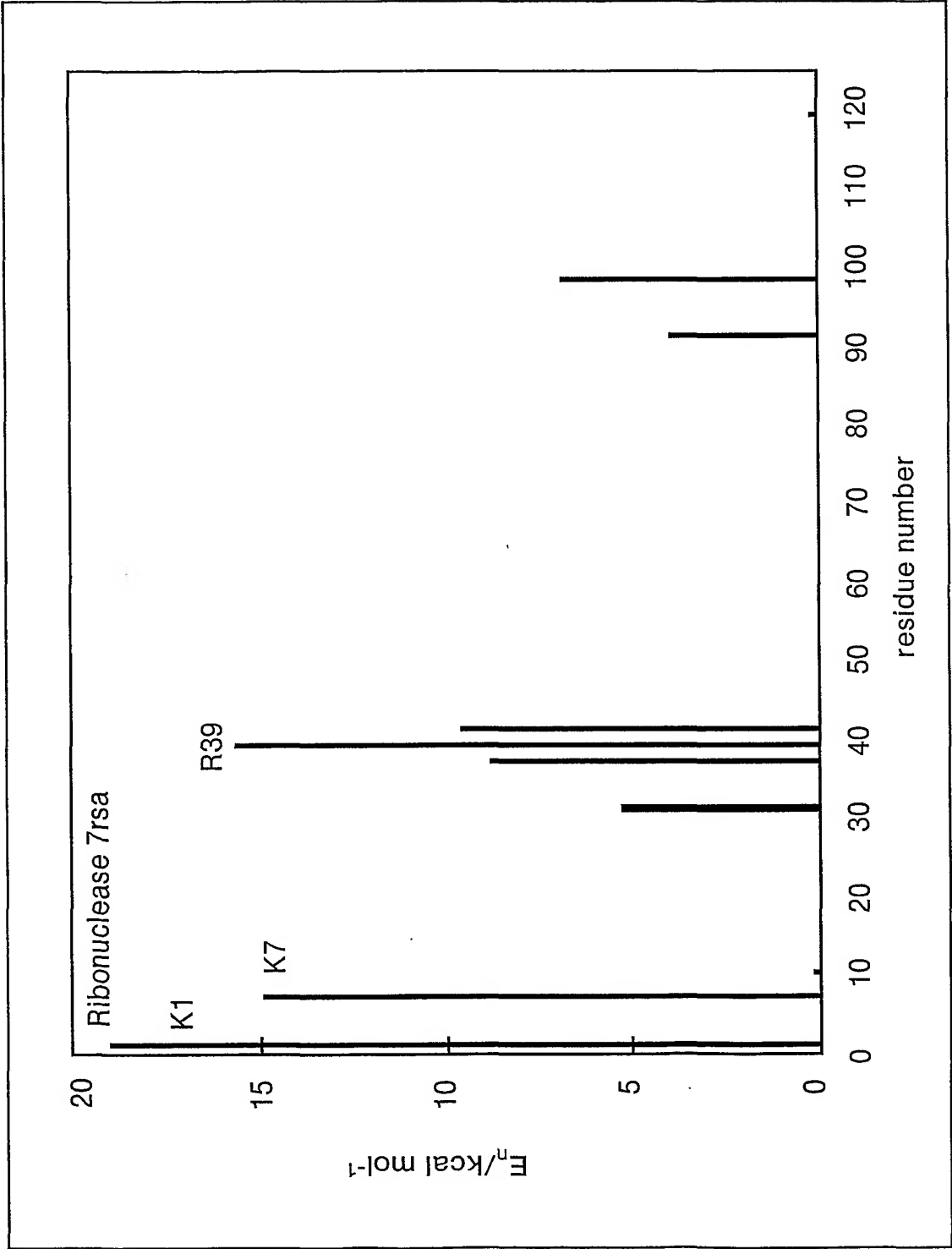




第46図

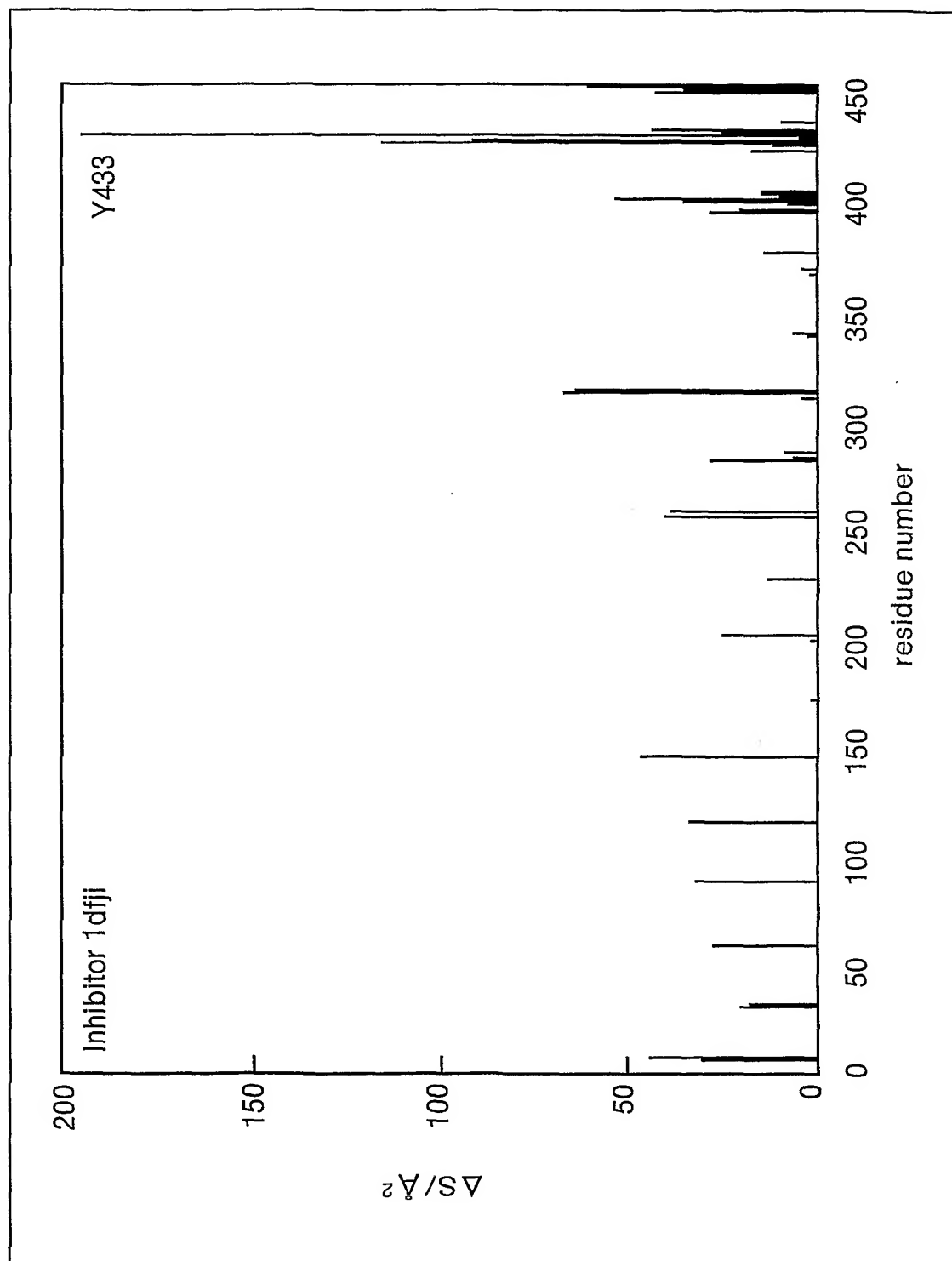


第47図

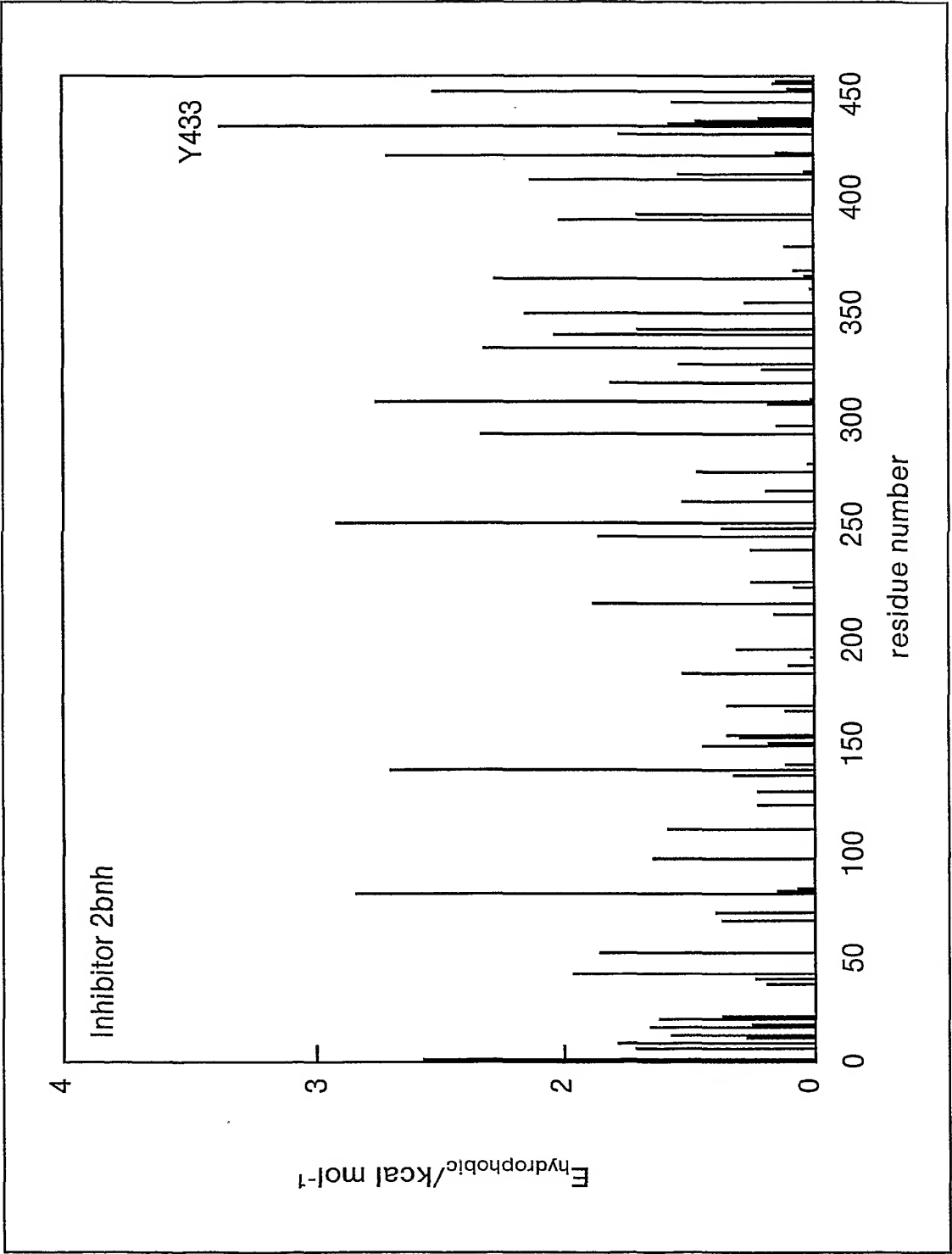


42/77

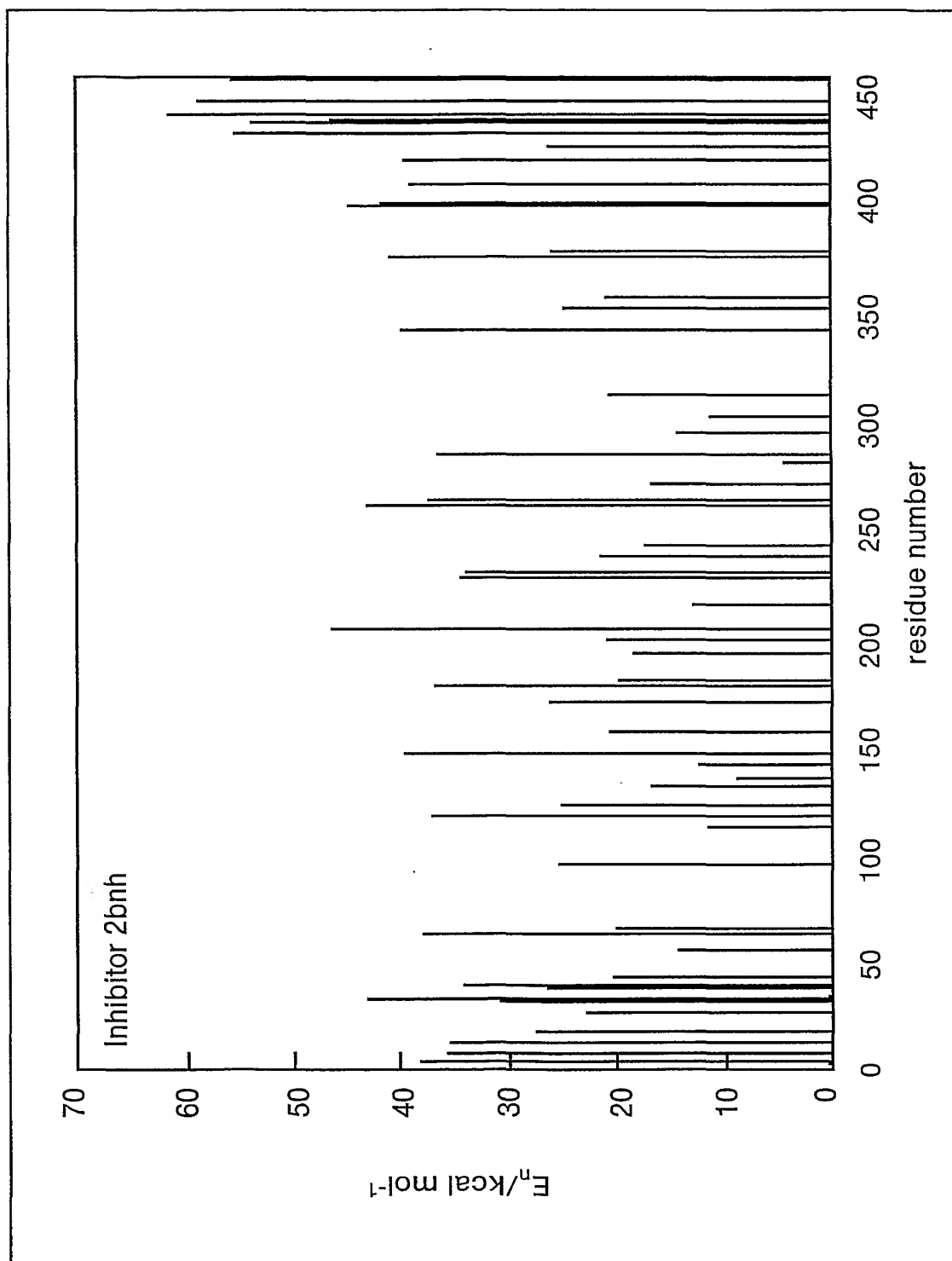
第48図



第49図



第50図



## 第51図

(本発明の基本原理1:1つの配列情報から蛋白質の結合部位を探す方法)

SA3-1  
 アミノ酸配列上の距離 $d$ や、立体構造予測シミュレーション結果や、  
 蛋白質構造DBから取得した立体構造座標データ等から、  
 蛋白質の立体構造の空間上の距離 $r$ を推定

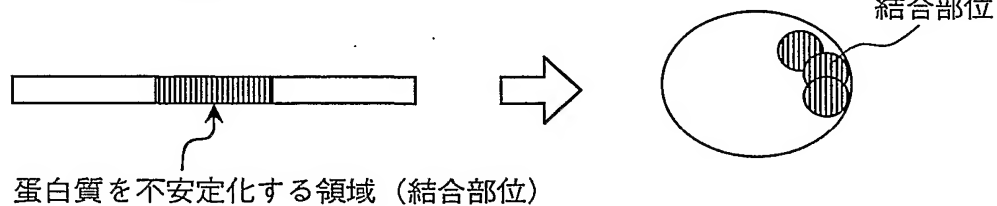
SA3-2  
 空間上の距離 $r$ から、蛋白質の全エネルギー $E_{total}$ を計算

$$\text{例) } E_{total} = 1/2 \sum \sum q_i q_j / r$$

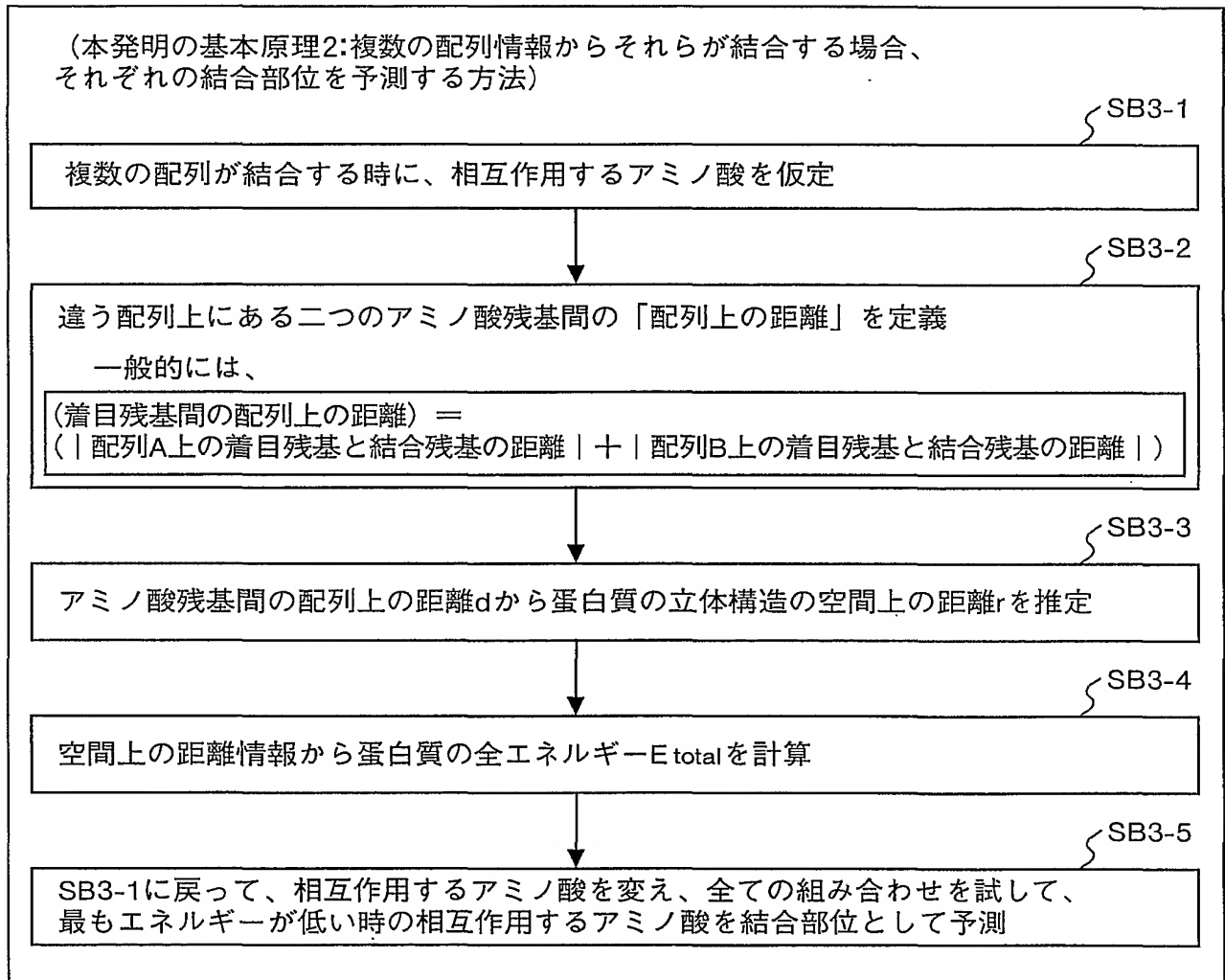
SA3-3  
 特定のアミノ酸と蛋白質内のそれ以外のアミノ酸残基との相互作用エネルギー  
 を以下の式により計算することにより、各アミノ酸残基が蛋白質の全エネルギー  
 をどれくらい安定化しているかを求める。

$$E_{interaction}(N) = q_N \sum q_j / r$$

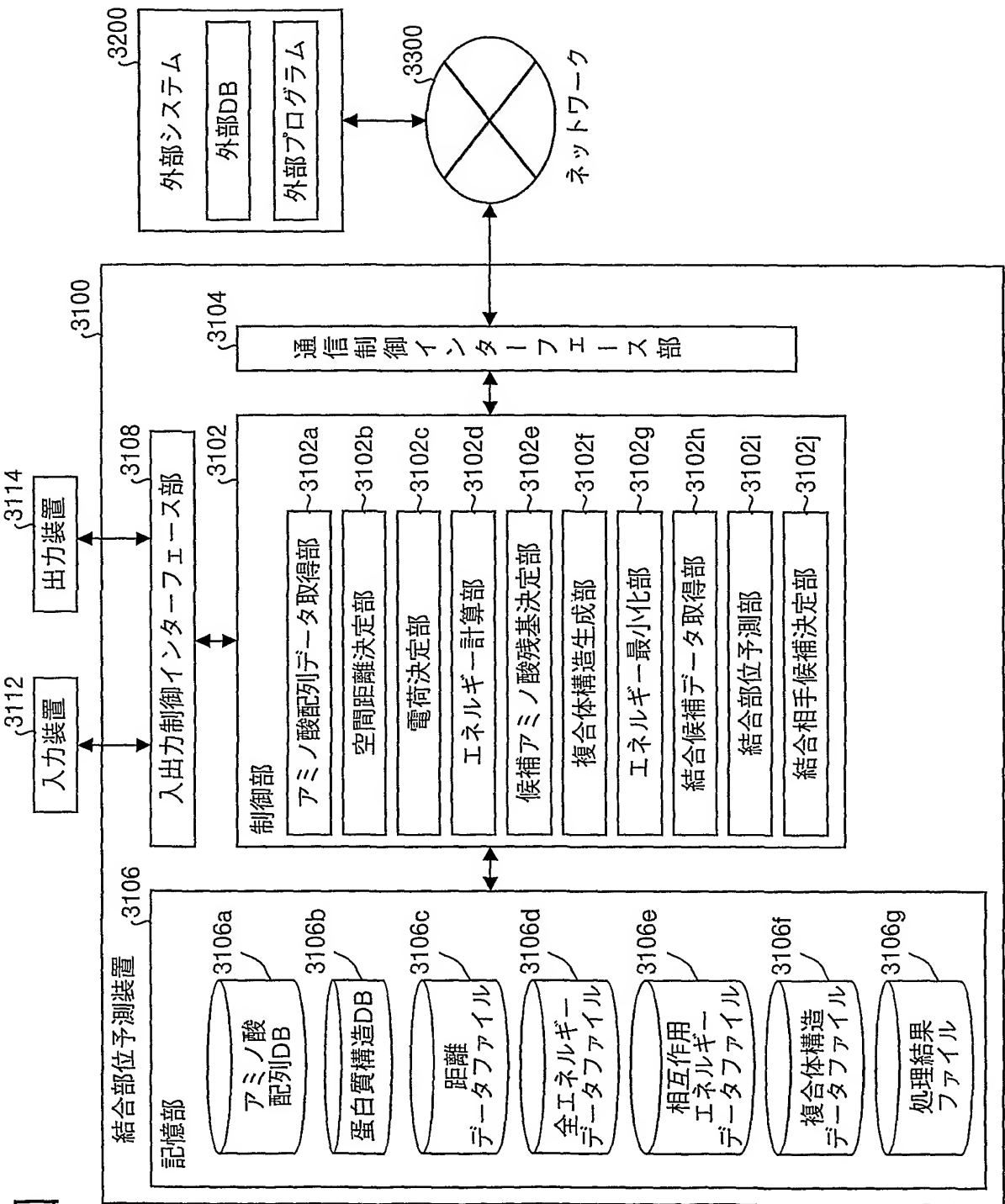
SA3-4  
 蛋白質をエネルギー的に不安定化させるアミノ酸残基がある領域を  
 結合部位の領域として予測



## 第52図

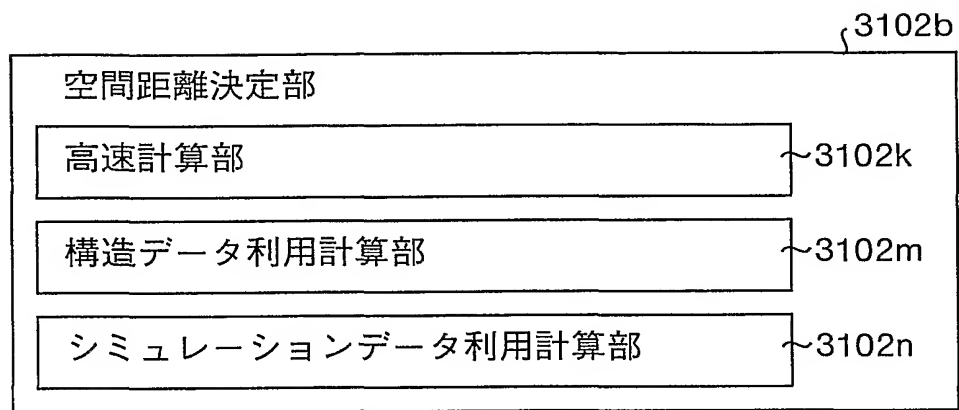


47/77

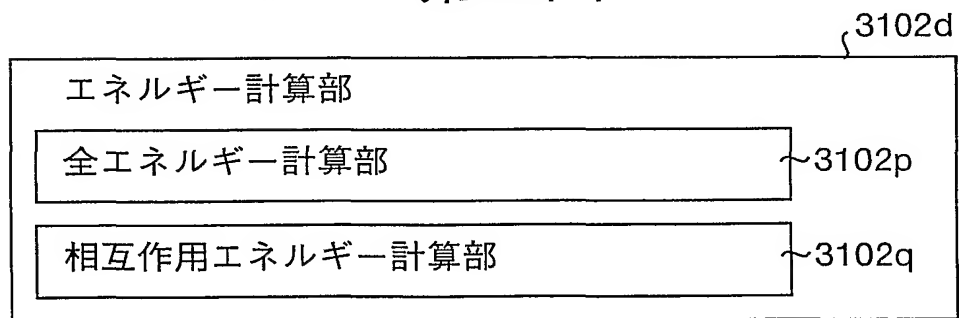




## 第54図

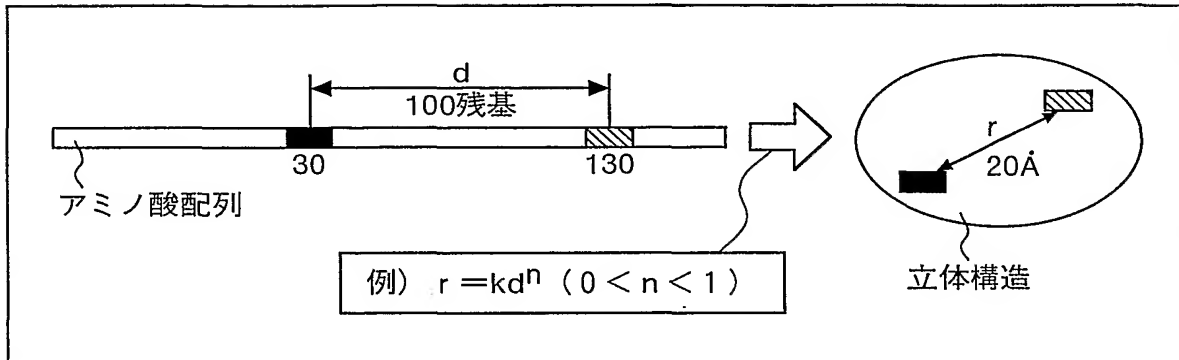


## 第55図

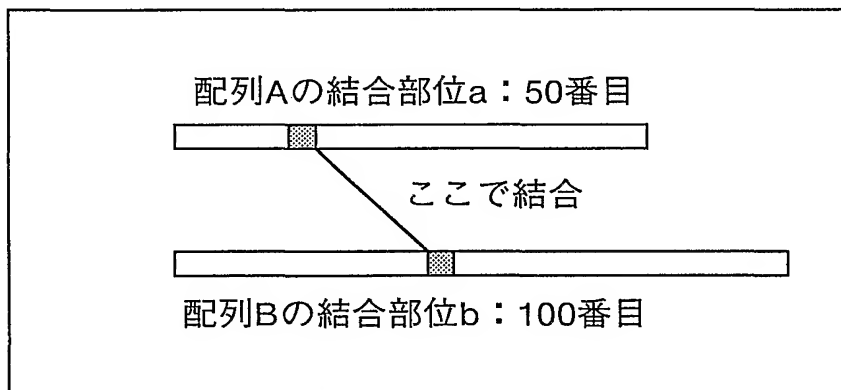


49/77

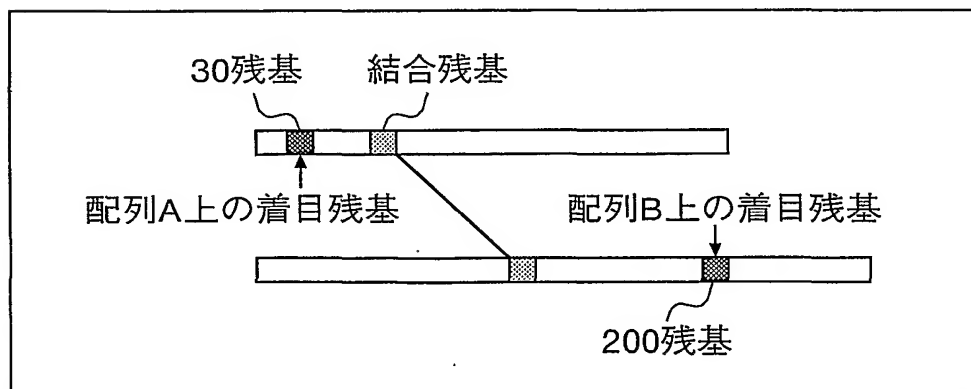
## 第56図



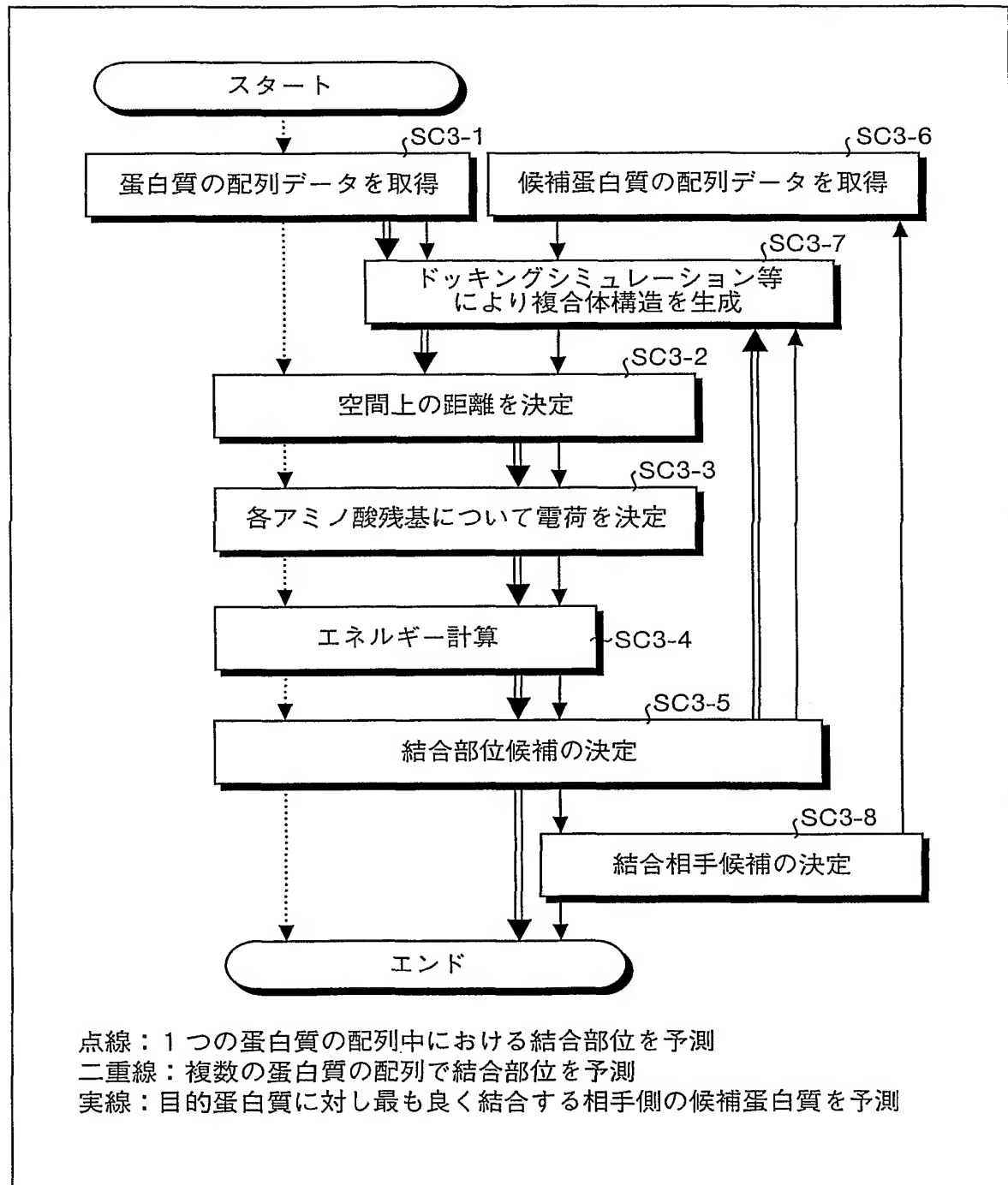
## 第57図



## 第58図



## 第59図

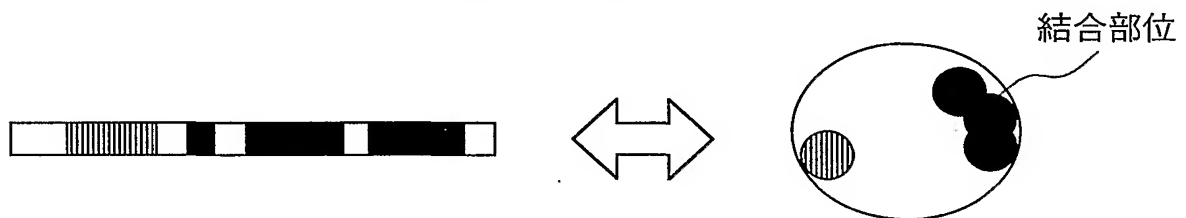


51/77

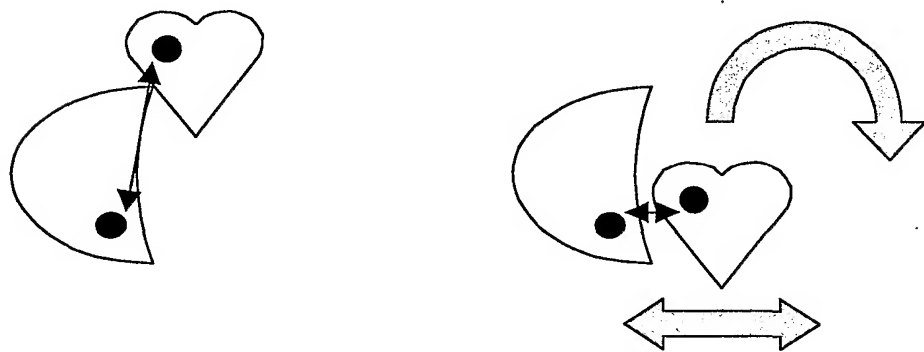
## 第60図

アミノ酸種類	アミノ酸番号	電荷	相互作用エネルギー
GLU	1	-1	0.14
ALA	2	0	0
ASP	3	-1	-0.34
ARG	4	1	-1.52
合計			-1.57

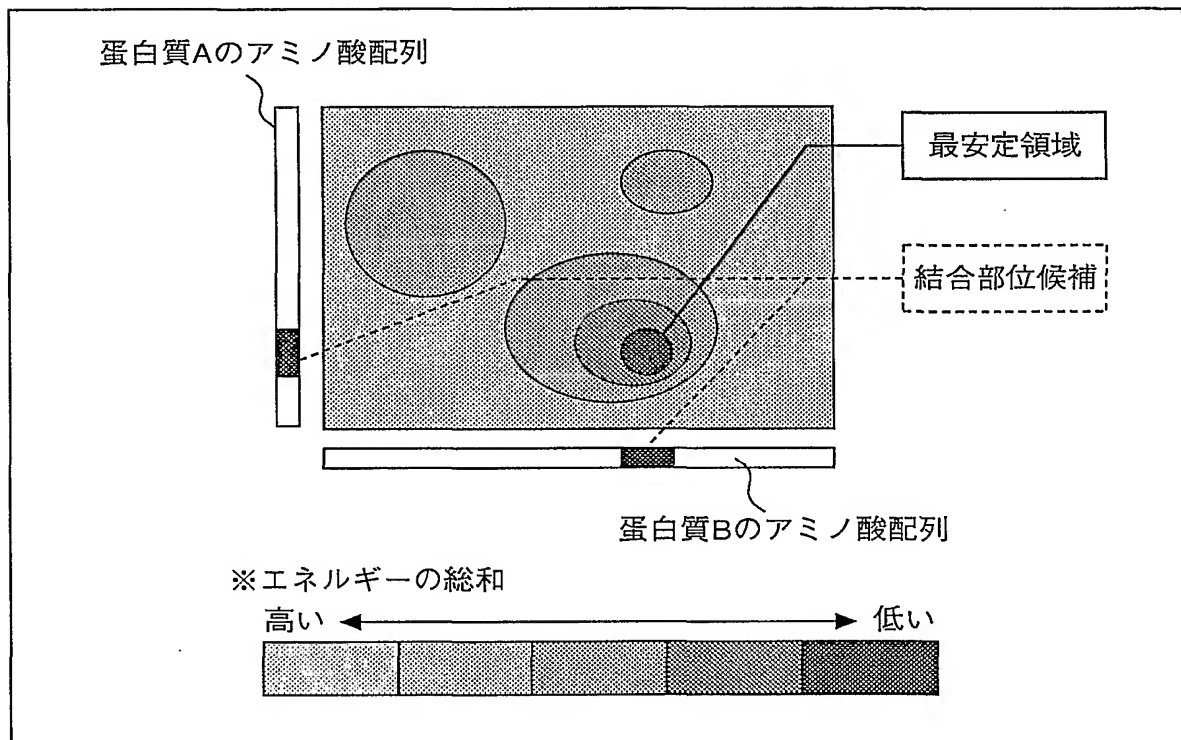
## 第61図



## 第62図

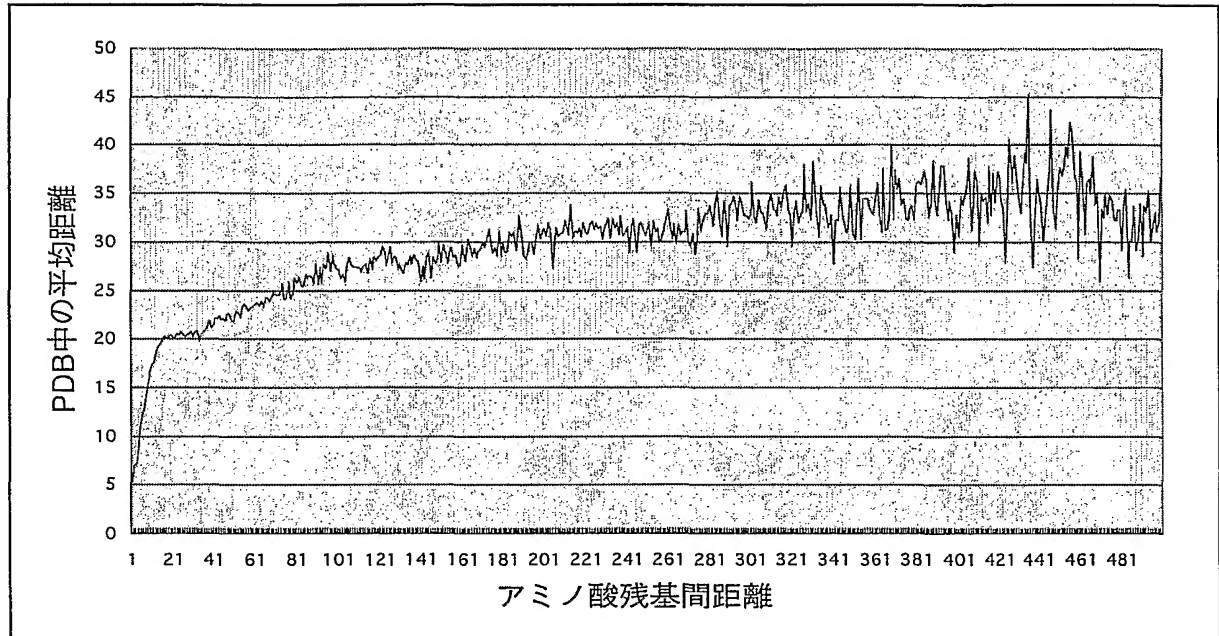


## 第63図

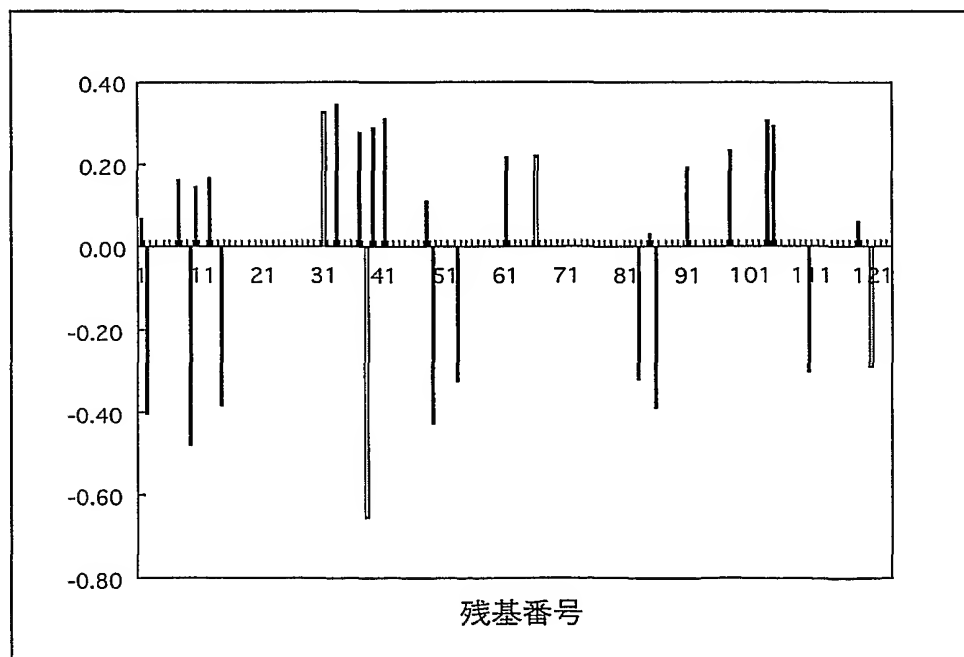


53/77

## 第64図



## 第65図



## 第66図

結合部位候補		
アミノ酸の種類	残基番号	
LYS	1	
LYS	7	結合部位
ARG	10	結合部位
HIS	12	結合部位
LYS	31	結合部位
ARG	33	
LYS	37	結合部位
ARG	39	結合部位
LYS	41	結合部位
HIS	48	
LYS	61	
LYS	66	結合部位
ARG	85	結合部位
LYS	91	結合部位
LYS	98	結合部位
LYS	104	
HIS	105	
HIS	119	結合部位

## 第67図

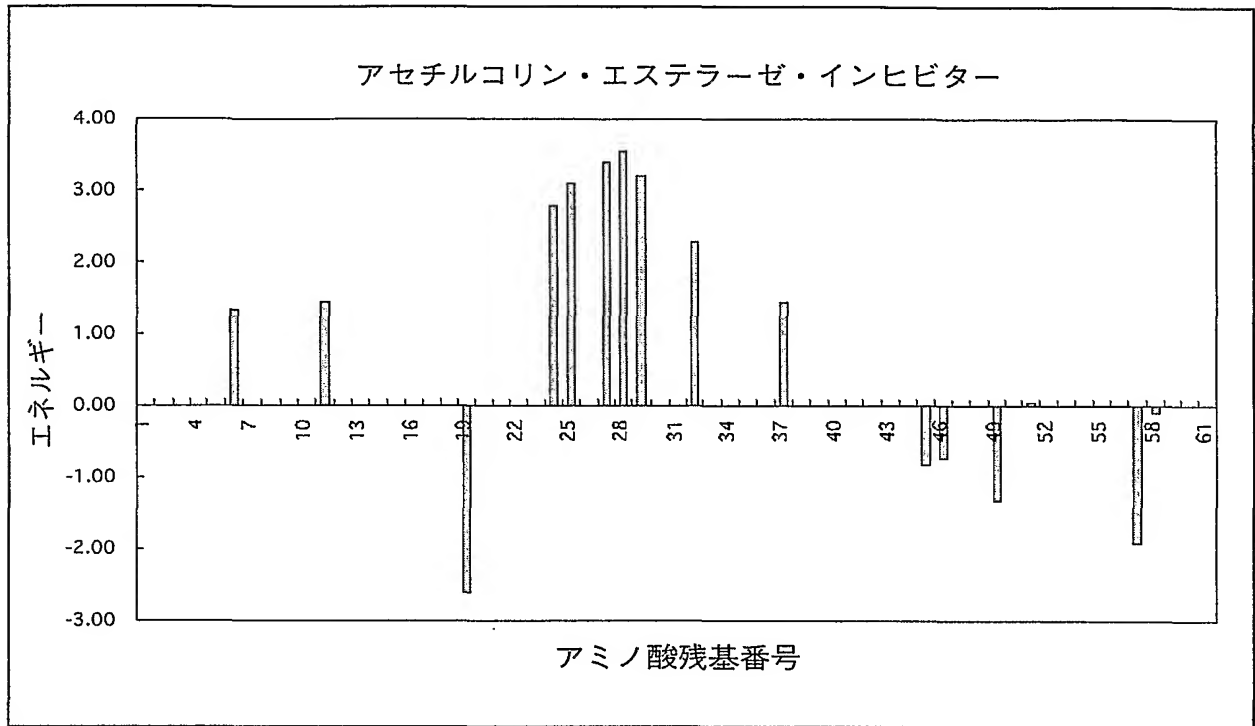
ATOM	1	N	GLU	A	4	31.787	8.114	206.704
ATOM	2	CA	GLU	A	4	32.664	8.451	205.542
ATOM	3	C	GLU	A	4	34.139	8.365	205.961
ATOM	4	O	GLU	A	4	34.597	9.206	206.741
ATOM	5	CB	GLU	A	4	32.362	9.878	205.041
ATOM	6	CG	GLU	A	4	31.206	10.023	204.032
ATOM	7	CD	GLU	A	4	29.82	9.97	204.657
ATOM	8	OE1	GLU	A	4	29.46	10.903	205.41
ATOM	9	OE2	GLU	A	4	29.081	9.003	204.373
ATOM	10	N	ASP	A	5	34.864	7.351	205.473
ATOM	11	CA	ASP	A	5	36.279	7.196	205.808
ATOM	12	C	ASP	A	5	37.048	5.998	205.209
ATOM	13	O	ASP	A	5	37.185	4.969	205.861
ATOM	14	CB	ASP	A	5	36.461	7.19	207.333
ATOM	15	CG	ASP	A	5	37.838	7.668	207.764
ATOM	16	OD1	ASP	A	5	38.692	7.942	206.889
ATOM	17	OD2	ASP	A	5	38.057	7.777	208.993
ATOM	18	N	PRO	A	6	37.449	6.084	203.93
ATOM	19	CA	PRO	A	6	38.206	5.043	203.22

※ 2列目から原子番号、原子種、chain名、アミノ酸残基番号、  
X座標、Y座標、Z座標を示す



56/77

## 第68図

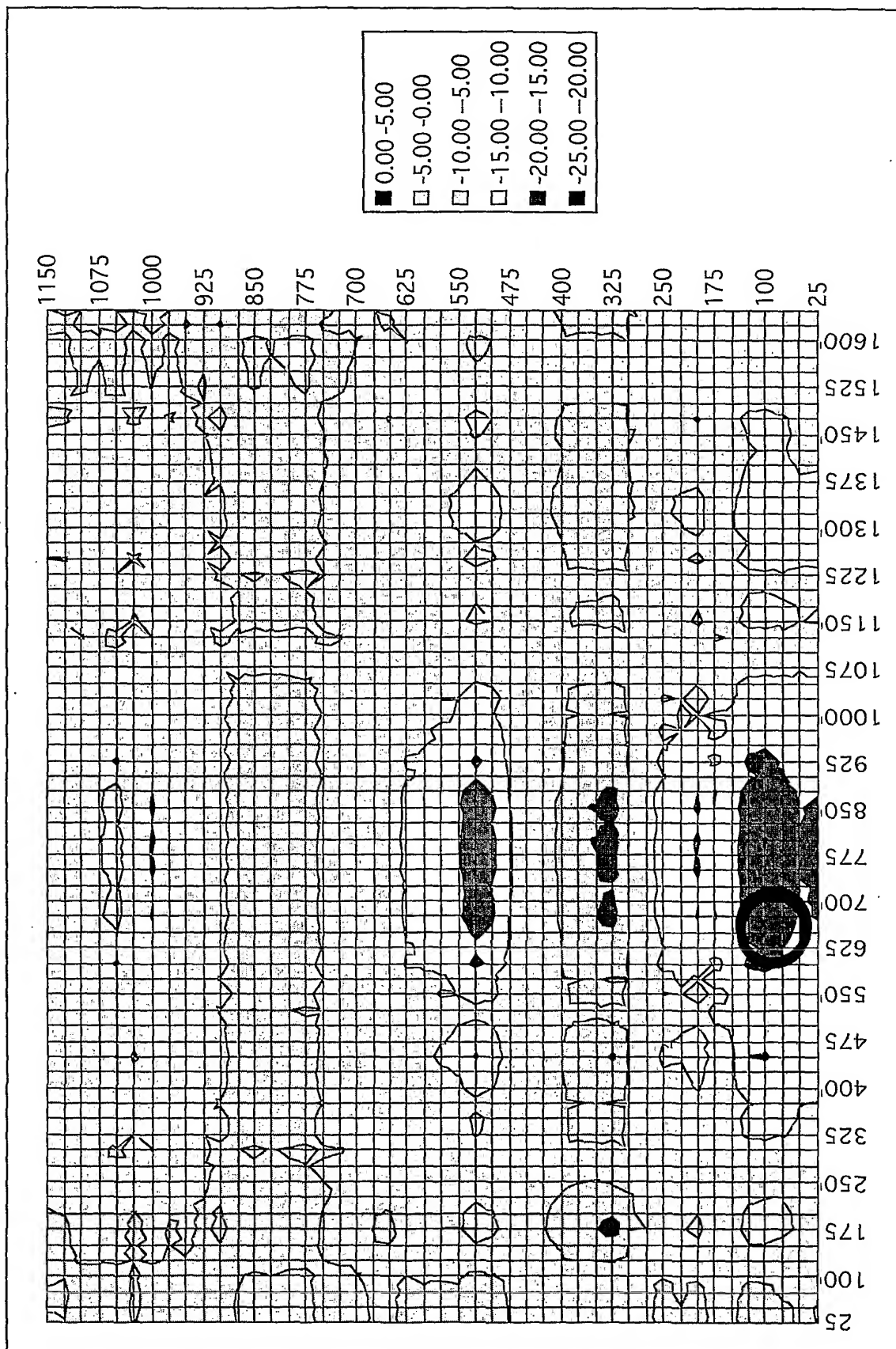


## 第69図

結合部位候補		
アミノ酸の種類	残基番号	
HIS	6	結合部位
ARG	11	結合部位
ARG	24	結合部位
LYS	25	
ARG	27	結合部位
ARG	28	
HIS	29	結合部位
LYS	32	結合部位
ARG	37	結合部位
LYS	51	

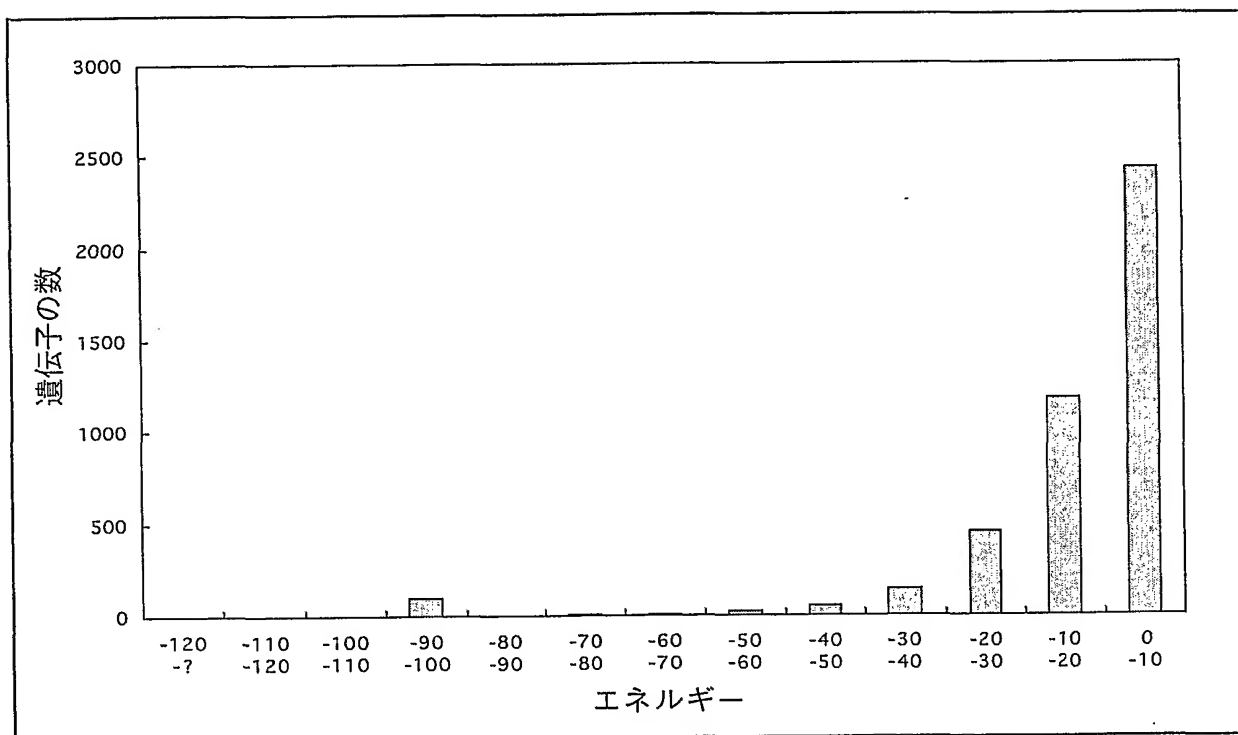
57/77

第70図



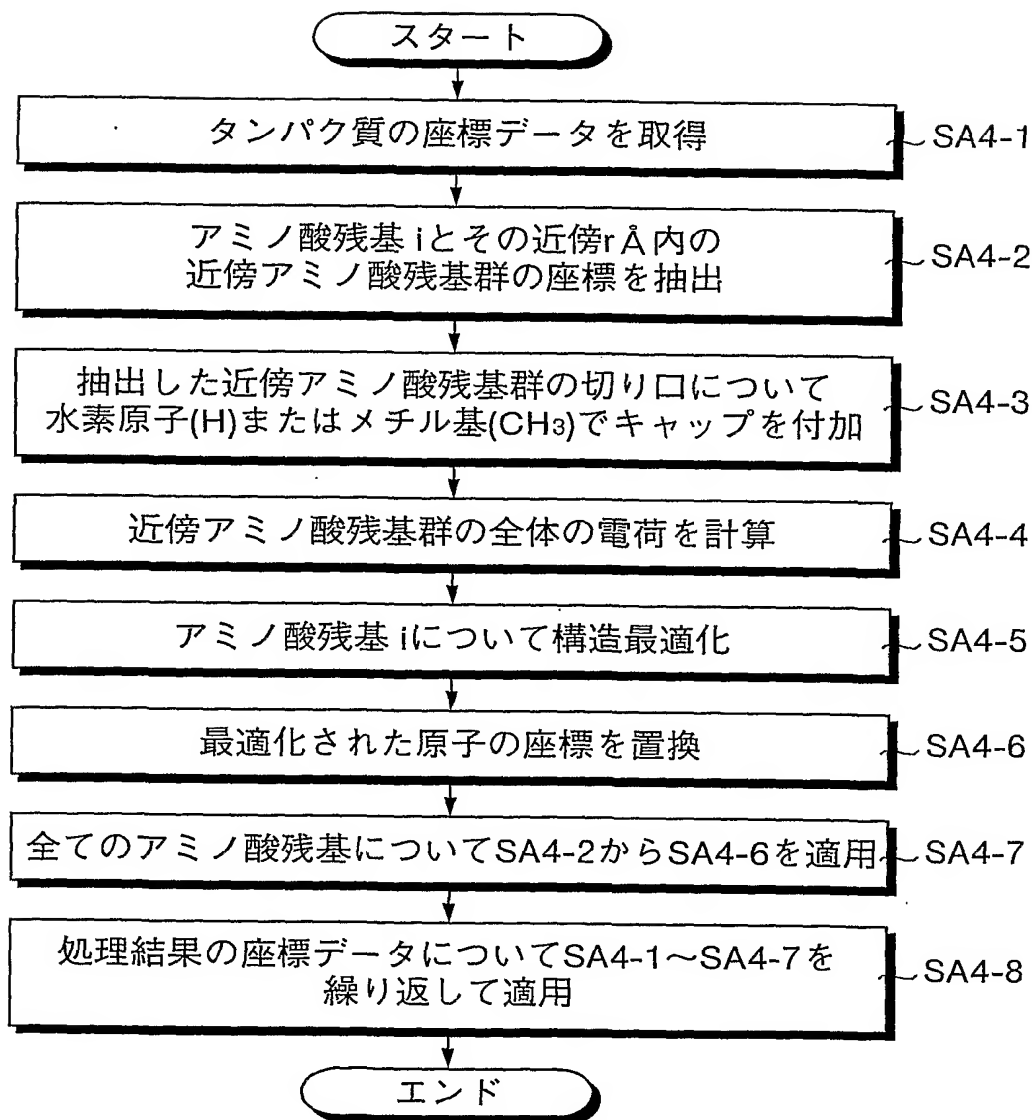
58/77

## 第71図

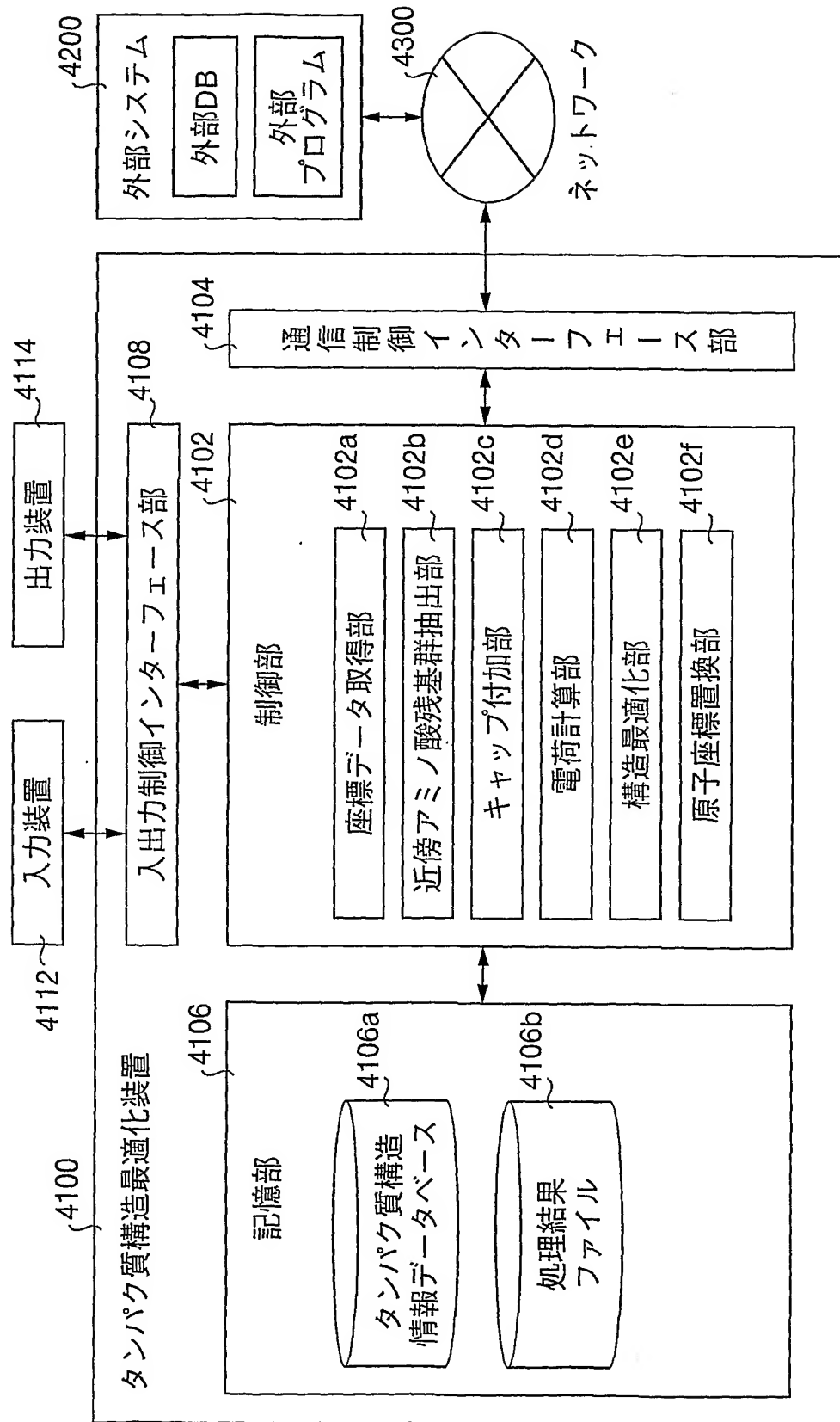


59/77

## 第72図

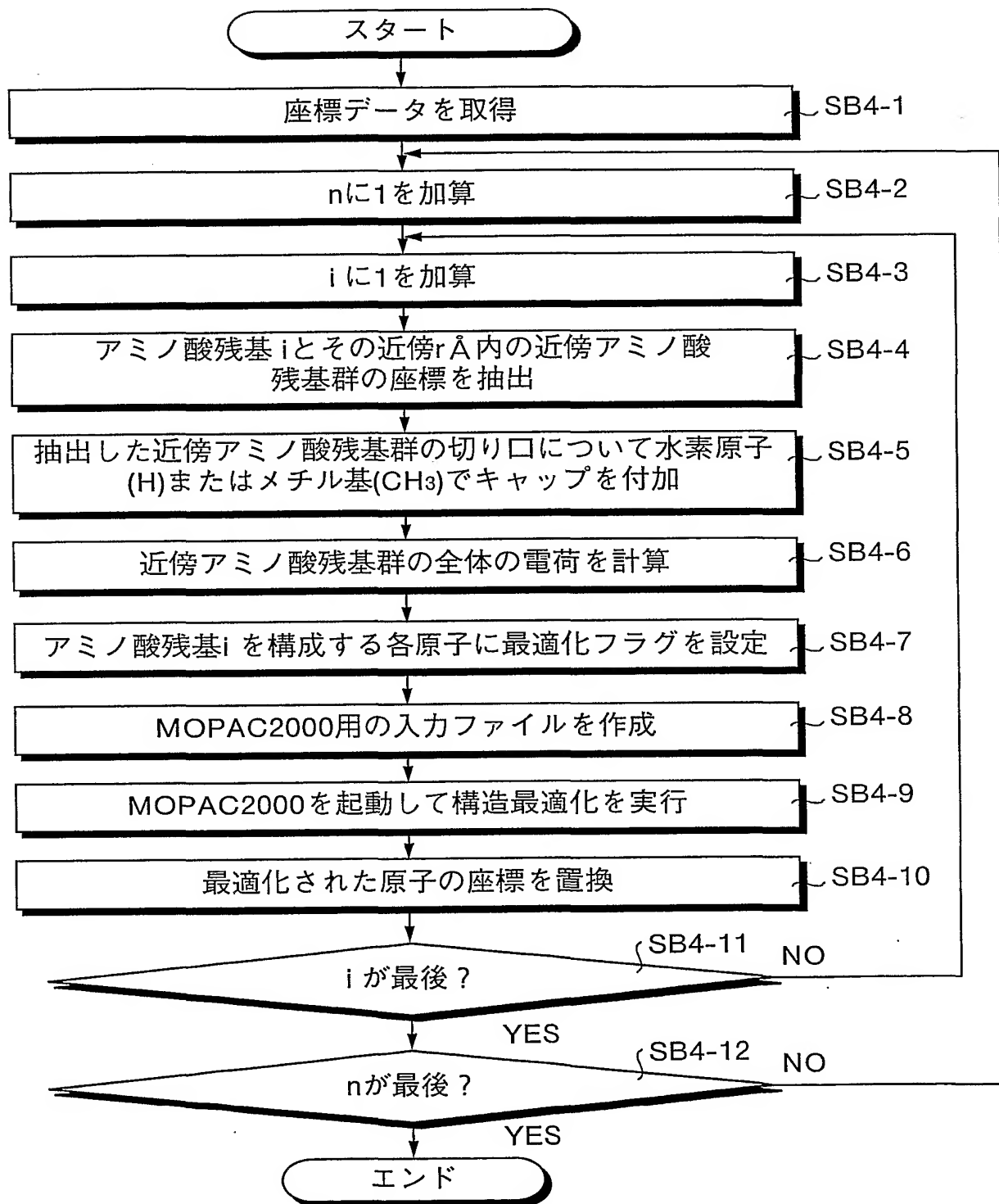


第73図



61/77

## 第74図



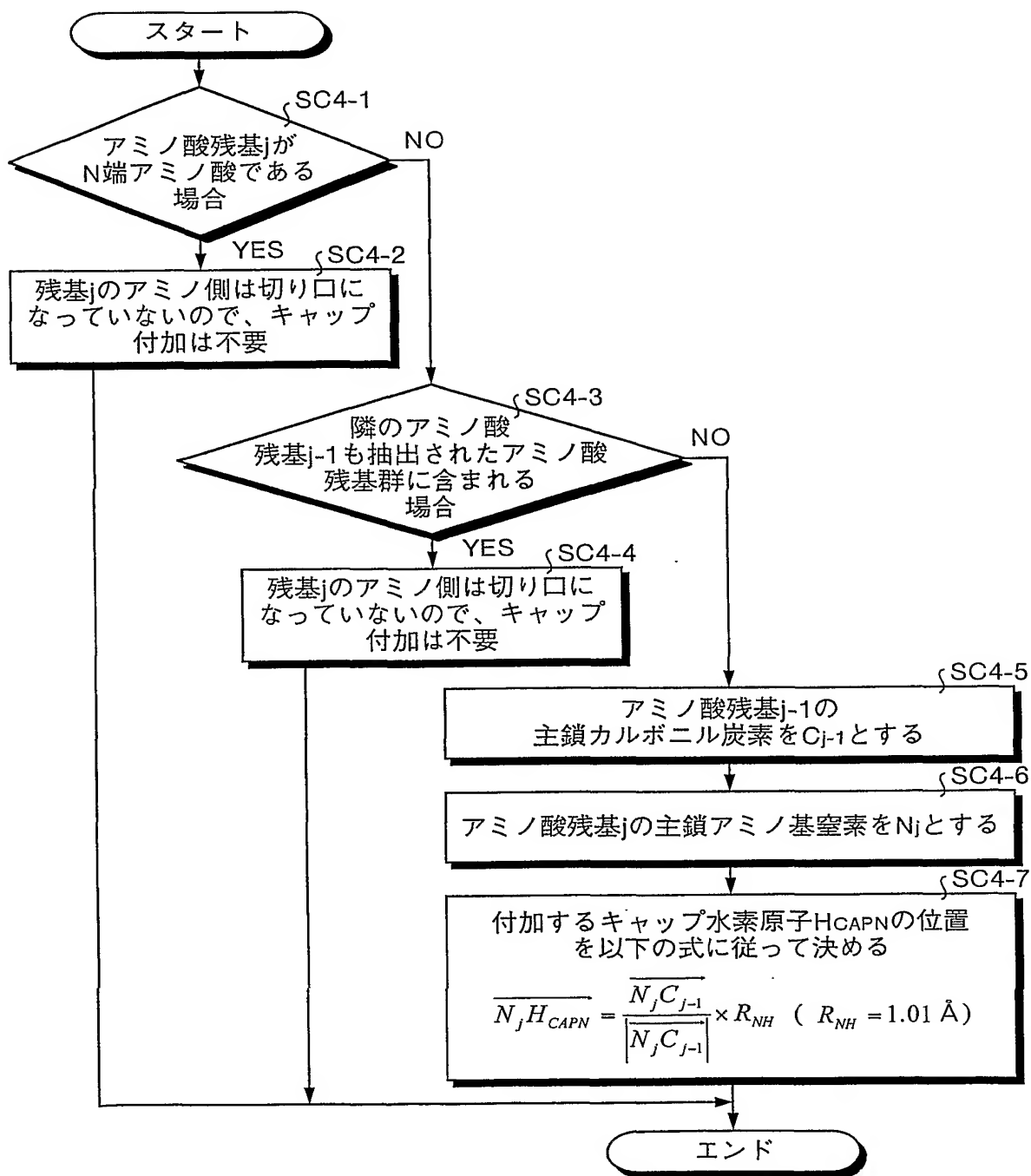
第75図

ATOM	1	N	PRO	1	1	7.854	104.656	32.974
ATOM	2	CA	PRO	1	1	7.699	104.034	31.626
ATOM	3	C	PRO	1	1	6.224	104.044	31.225
ATOM	4	O	PRO	1	1	5.374	104.542	31.976
ATOM	5	CB	PRO	1	1	8.252	102.610	31.704
ATOM	6	CG	PRO	1	1	8.313	102.380	33.238
ATOM	7	CD	PRO	1	1	8.637	103.754	33.838
ATOM	8	H1	PRO	1	1	8.307	105.545	32.880
ATOM	9	H1 9	PRO	1	1	8.262	104.633	30.852
ATOM	10	H2 1	PRO	1	1	7.604	101.848	31.214
ATOM	11	H2 3	PRO	1	1	9.272	102.555	31.254
ATOM	12	H2 5	PRO	1	1	7.326	102.006	33.609
ATOM	13	H2 7	PRO	1	1	9.087	101.621	33.506
ATOM	14	H2 9	PRO	1	1	8.268	103.859	34.890
ATOM	15	H3 1	PRO	1	1	9.740	103.958	33.814
ATOM	16	N	ASN	1	2	5.932	103.521	30.041
ATOM	17	CA	ASN	1	2	4.573	103.458	29.522
ATOM	18	C	ASN	1	2	4.220	101.980	29.427
ATOM	19	O	ASN	1	2	4.928	101.148	29.993
ATOM	20	CB	ASN	1	2	4.520	104.105	28.139
ATOM	21	CG	ASN	1	2	5.450	103.428	27.141
ATOM	22	OD 1	ASN	1	2	6.047	102.391	27.440

## 途中省略

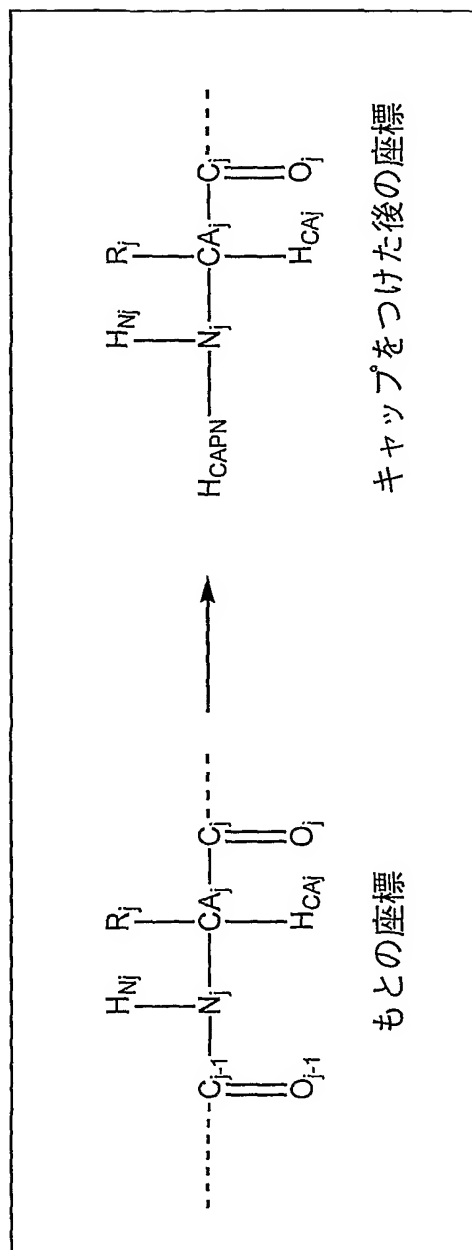
ATOM	2132	H40	ARG	1	135	7.570	94.945	31.002
ATOM	2133	H42	ARG	1	135	7.667	95.518	32.710
ATOM	2134	H44	ARG	1	135	9.663	96.855	32.278
ATOM	2135	H46	ARG	1	135	10.017	95.974	30.735
ATOM	2136	N	GLU	1	136	9.286	91.036	30.707
ATOM	2137	CA	GLU	1	136	10.105	89.902	30.320
ATOM	2138	C	GLU	1	136	11.401	90.443	29.726
ATOM	2139	O	GLU	1	136	12.410	89.708	29.745
ATOM	2140	CB	GLU	1	136	9.331	89.094	29.277
ATOM	2141	CG	GLU	1	136	10.053	87.930	28.638
ATOM	2142	CD	GLU	1	136	9.277	87.369	27.463
ATOM	2143	OE1	GLU	1	136	8.096	87.002	27.650
ATOM	2144	OE2	GLU	1	136	9.838	87.319	26.346
ATOM	2145	OXT	GLU	1	136	11.392	91.606	29.258
ATOM	2146	H	GLU	1	136	8.772	91.468	29.962
ATOM	2147	H22	GLU	1	136	10.370	89.273	31.218
ATOM	2148	H24	GLU	1	136	8.395	88.698	29.761
ATOM	2149	H26	GLU	1	136	9.002	89.782	28.445
ATOM	2150	H28	GLU	1	136	11.071	88.252	28.293
ATOM	2151	H30	GLU	1	136	10.198	87.111	29.390
ATOM	2152	H32	GLU	1	136	12.254	91.868	28.859
TER								

## 第76図

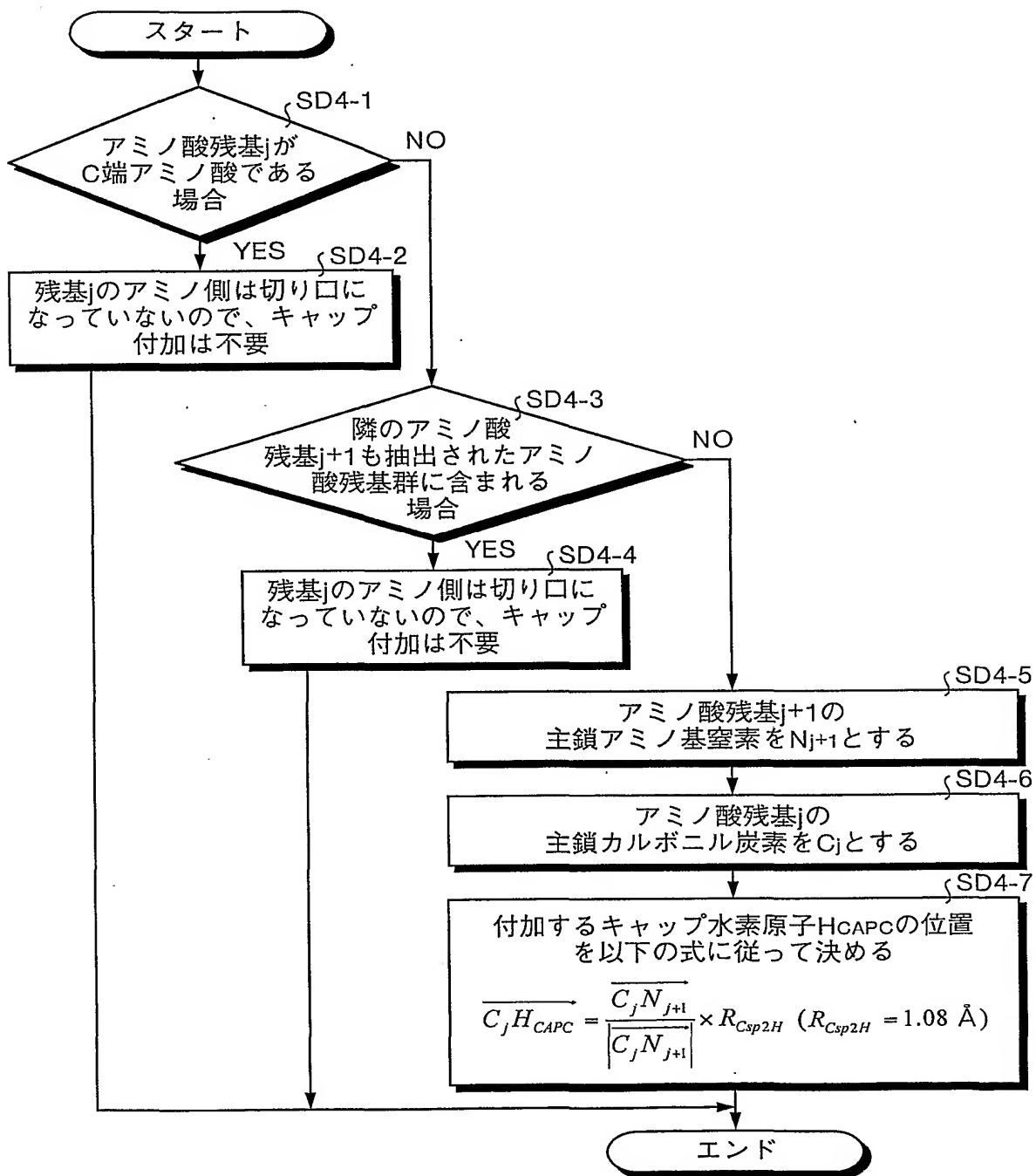




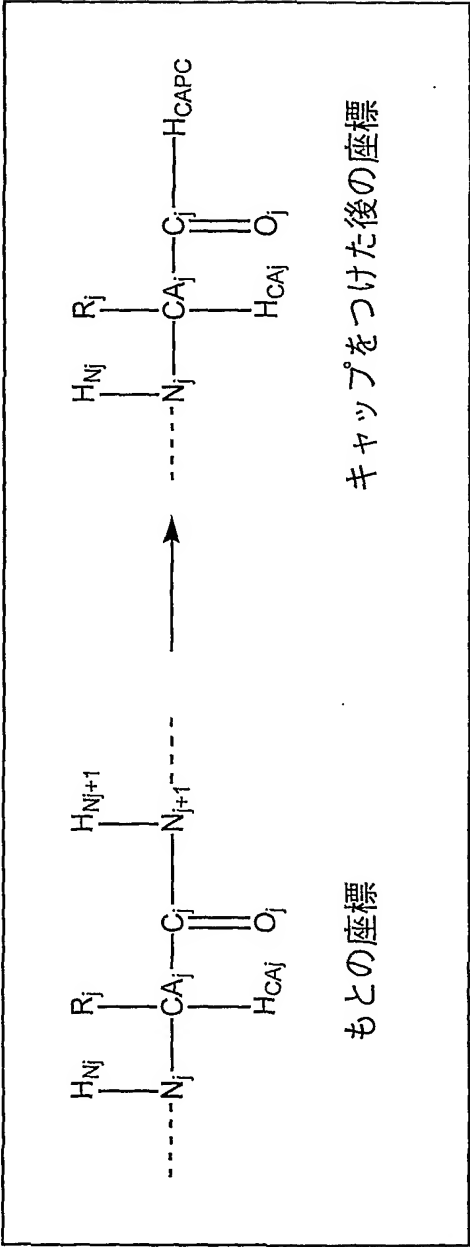
第77図



## 第78図

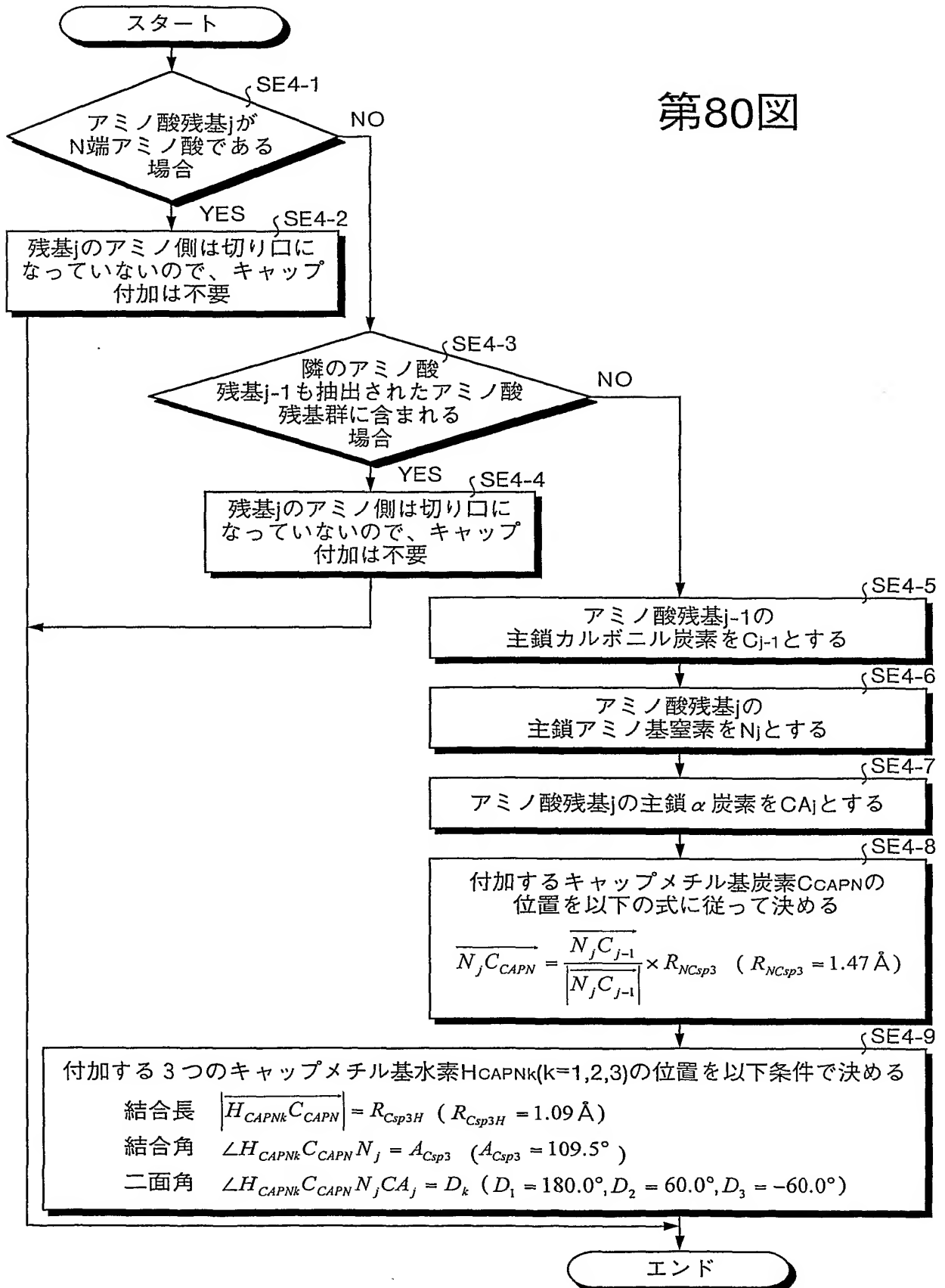


第79図

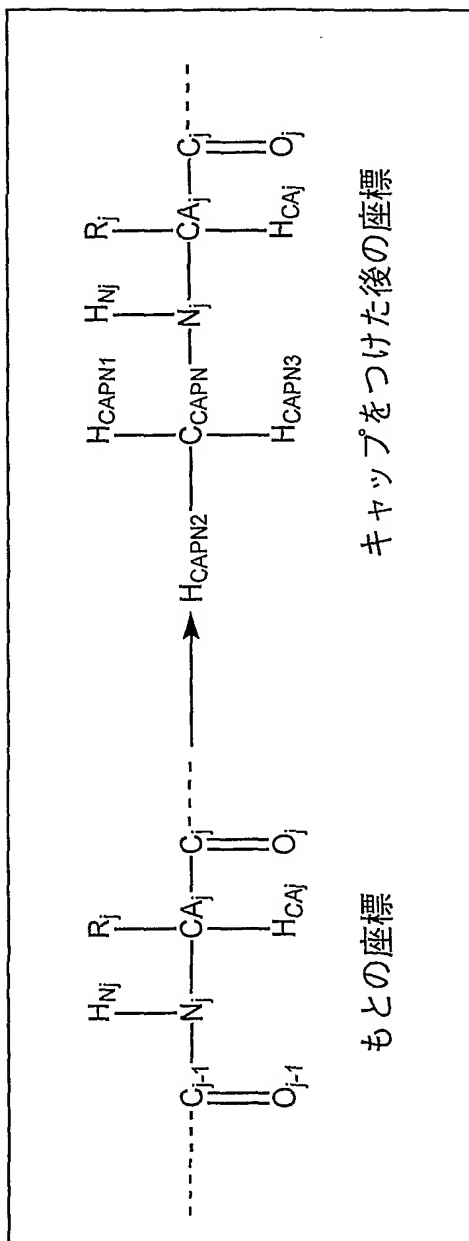


67/77

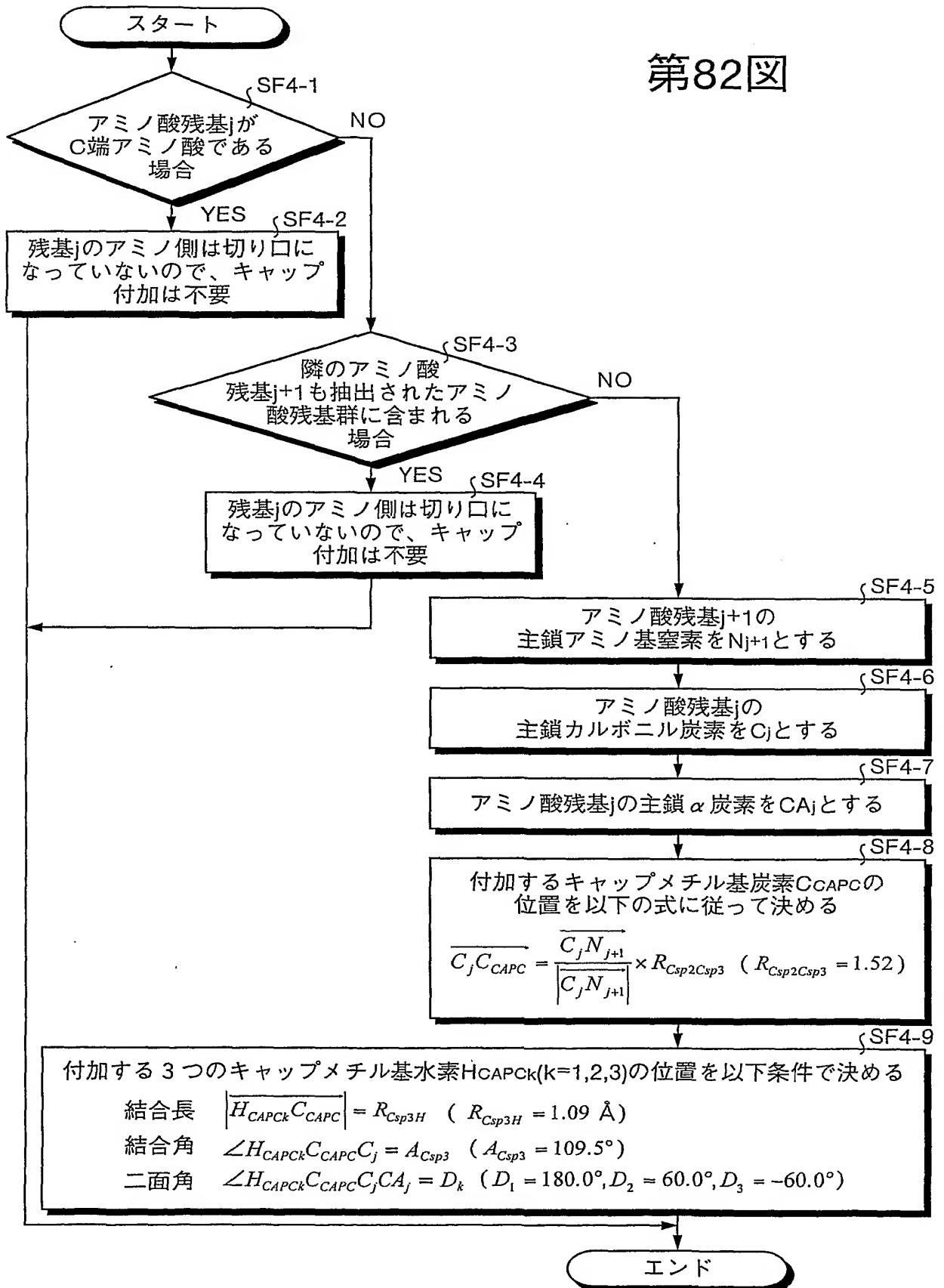
## 第80図



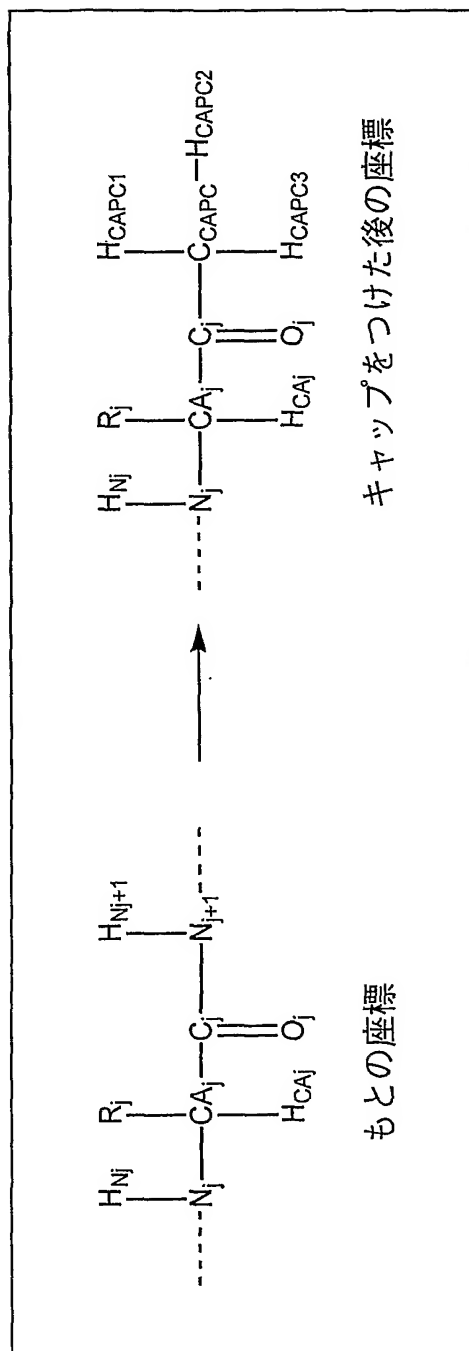
## 第81図



## 第82図



目錄 38



## 第84図

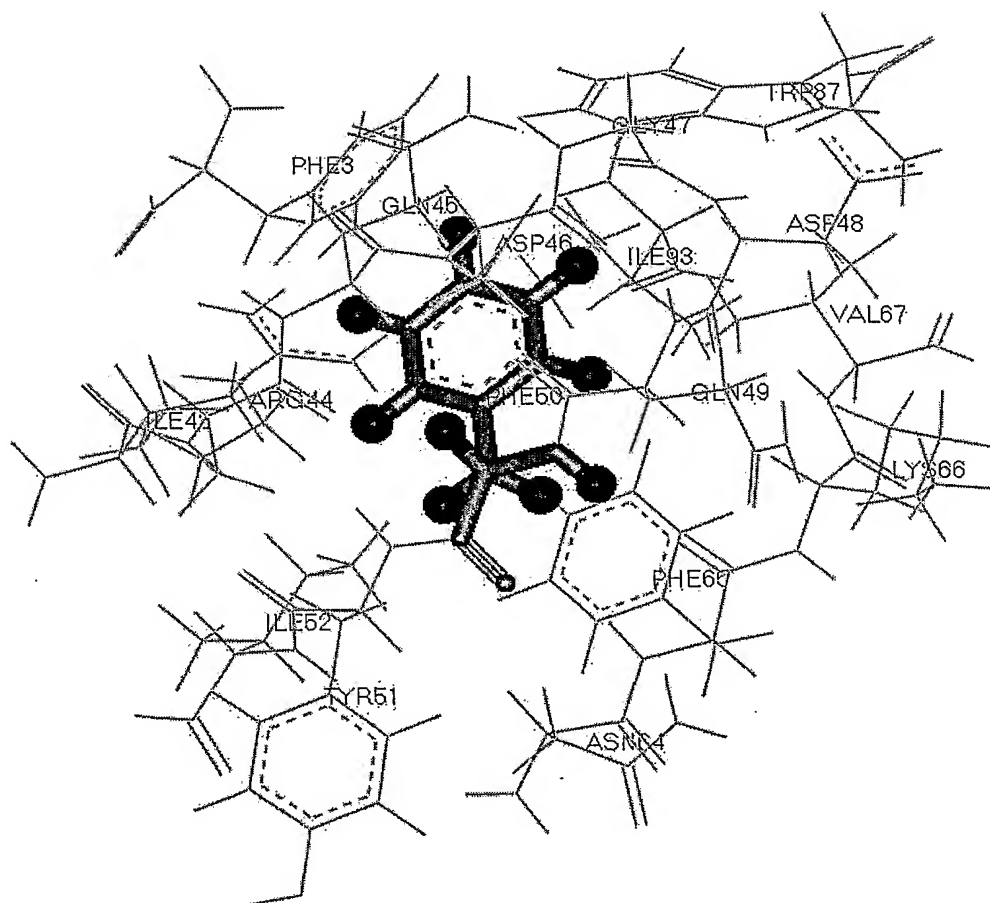
この3文字でアミノ酸の種類を判別する

.....  
ATOM 2132 H40 ARG 1 135 7.570 94.945 31.002  
ATOM 2133 H42 ARG 1 135 7.667 95.518 32.710  
ATOM 2134 H44 ARG 1 135 9.663 96.855 32.278  
ATOM 2135 H46 ARG 1 135 10.017 95.974 30.735  
.....



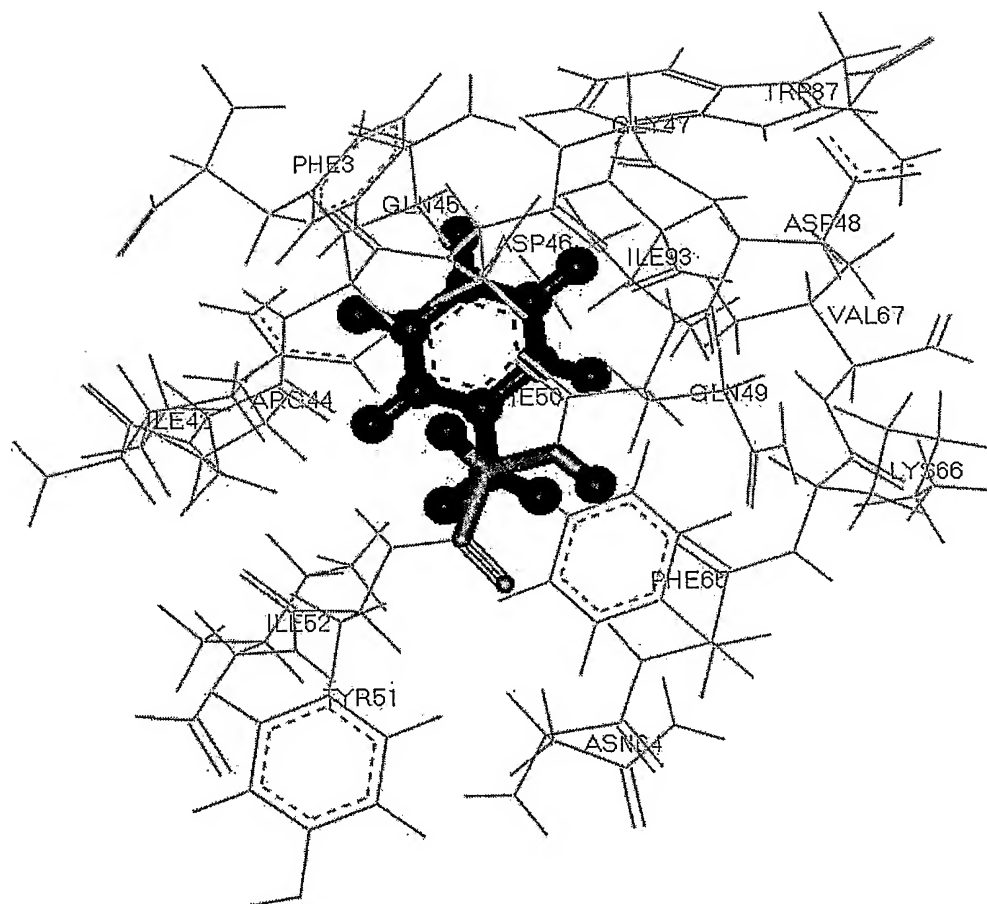
72/77

## 第85図



73/77

## 第86図



74/77

## 第87図

最適化フラグ この値が1の場合最適化し、  
0の場合には最適化しない

計算した電荷の値をここに入れる

原子種      カッコの内側はコメント

---ここから---  
am1 mmok mmccrok cycles=8000 eps=4.0 memory=2000 &  
& charge=0  
1cbl F50

x,y,z座標

キーワード行  
キーワード行  
コメント行  
以下初期座標データ

N(	1	PHE	3	N )	3.17000	0	101.65400	0	28.67700	0
C(	1	PHE	3	CA )	2.71300	0	100.27000	0	28.51000	0
C(	1	PHE	3	C )	3.50800	0	99.45700	0	27.50100	0
O(	1	PHE	3	O )	3.34800	0	98.23900	0	27.43500	0
C(	1	PHE	3	CB )	1.25400	0	100.23300	0	28.06200	0
C(	1	PHE	3	CG )	0.27400	0	100.69400	0	29.10000	0
C(	1	PHE	3	CD1)	0.17900	0	102.03200	0	29.44600	0
C(	1	PHE	3	CD2)	-0.60100	0	99.79500	0	29.68400	0
C(	1	PHE	3	CE1)	-0.77400	0	102.46500	0	30.35200	0
C(	1	PHE	3	CE2)	-1.56100	0	100.22600	0	30.59400	0
C(	1	PHE	3	CZ )	-1.64700	0	101.56300	0	30.92600	0
H(	1	PHE	3	H )	2.57000	0	102.35900	0	28.32400	0
.....										
途中省略										
.....										
H(	1	GLN	49	HE2)	-2.22300	0	107.41300	0	15.77000	0
N(	1	PHE	50	N )	-2.36600	0	102.06400	0	20.36900	0
C(	1	PHE	50	CA )	-1.82400	0	100.79200	0	20.84100	0
C(	1	PHE	50	C )	-1.88200	0	99.70400	0	19.79200	0
O(	1	PHE	50	O )	-2.80700	0	99.65400	0	18.98500	0
C(	1	PHE	50	CB )	-2.65100	0	100.25800	0	22.01700	0
C(	1	PHE	50	CG )	-2.29600	0	100.85000	0	23.32600	0
C(	1	PHE	50	CD1)	-2.69800	0	102.13800	0	23.65300	0
C(	1	PHE	50	CD2)	-1.55000	0	100.12300	0	24.23900	0
C(	1	PHE	50	CE1)	-2.35500	0	102.69200	0	24.87100	0
C(	1	PHE	50	CE2)	-1.20300	0	100.66700	0	25.45700	0
C(	1	PHE	50	CZ )	-1.60200	0	101.94900	0	25.77600	0
H(	1	PHE	50	H )	-3.28500	1	102.01600	1	19.99400	1
H(	1	PHE	50	HA )	-0.75800	1	100.98600	1	21.18400	1
H(	1	PHE	50	HB )	-3.74900	1	100.42200	1	21.82100	1
H(	1	PHE	50	HB )	-2.50800	1	99.14100	1	22.07700	1
H(	1	PHE	50	HD1)	-3.30600	1	102.72700	1	22.94200	1
H(	1	PHE	50	HD2)	-1.23400	1	99.09400	1	23.99300	1
.....										
途中省略										
.....										

## 第88図

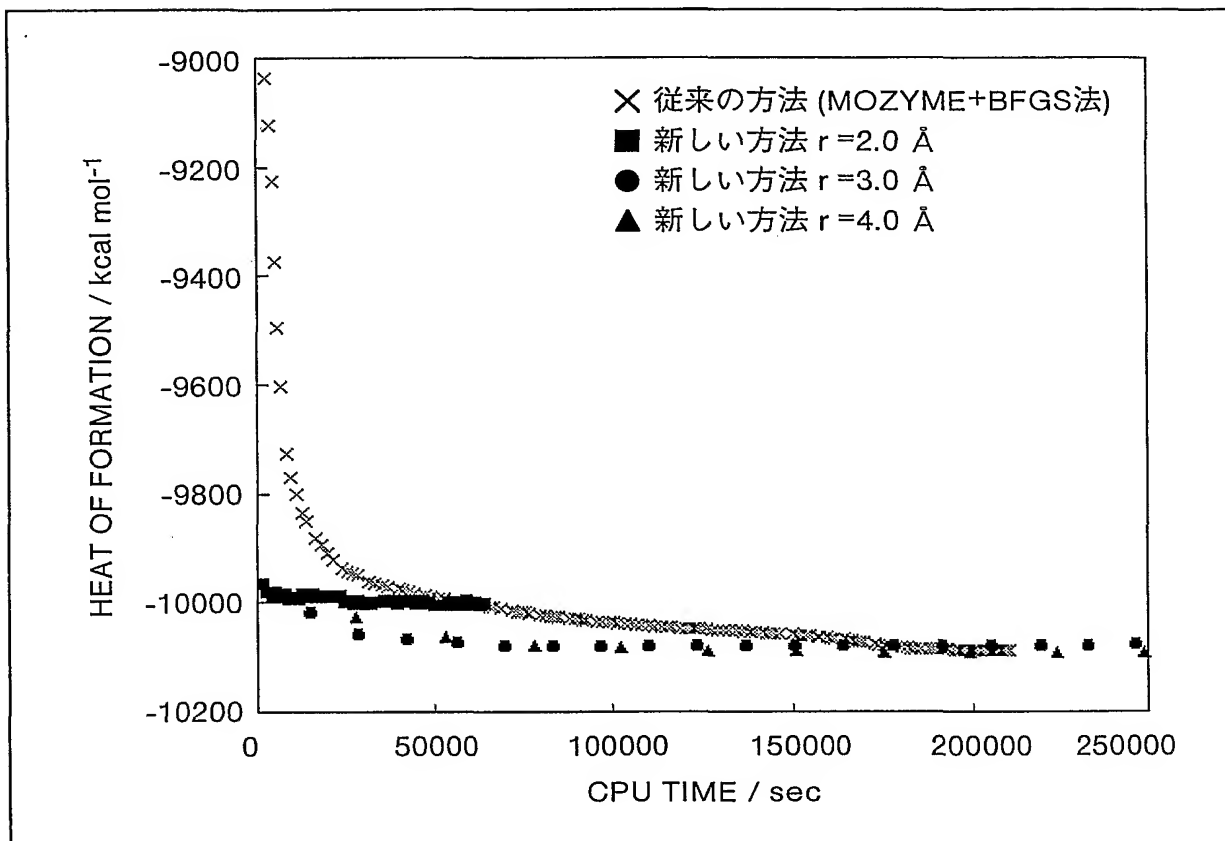
途省略

131 H (1 GLN 49 H) -2. 223000000	107. 413000000	15. 770000000	*
132 N (1 PHE 50 N) -2. 366000000	102. 064000000	20. 369000000	*
133 C (1 PHE 50 C) -1. 824000000	100. 792000000	20. 841000000	*
134 C (1 PHE 50 C) -1. 882000000	99. 704000000	19. 792000000	*
135 O (1 PHE 50 O) -2. 807000000	99. 654000000	18. 985000000	*
136 C (1 PHE 50 C) -2. 651000000	100. 258000000	22. 017000000	*
137 C (1 PHE 50 C) -2. 296000000	100. 850000000	23. 326000000	*
138 C (1 PHE 50 C) -2. 698000000	102. 138000000	23. 653000000	*
139 C (1 PHE 50 C) -1. 550000000	100. 123000000	24. 239000000	*
140 C (1 PHE 50 C) -2. 355000000	102. 692000000	24. 871000000	*
141 C (1 PHE 50 C) -1. 203000000	100. 667000000	25. 457000000	*
142 C (1 PHE 50 C) -1. 602000000	101. 949000000	25. 776000000	*
143 H (1 PHE 50 H) -3. 285000000	102. 016000000	19. 994000000	*
144 H (1 PHE 50 H) -0. 758000000	100. 986000000	21. 184000000	*
145 H (1 PHE 50 H) -3. 749000000	100. 422000000	21. 821000000	*
146 H (1 PHE 50 H) -2. 508000000	99. 141000000	22. 077000000	*
147 H (1 PHE 50 H) -3. 306000000	102. 727000000	22. 942000000	*
148 H (1 PHE 50 H) -1. 234000000	99. 094000000	23. 993000000	*
149 H (1 PHE 50 H) -2. 682000000	103. 712000000	25. 129000000	*
150 H (1 PHE 50 H) -0. 614000000	100. 074000000	26. 178000000	*
151 H (1 PHE 50 H) -1. 318000000	102. 389000000	26. 747000000	*
152 N (1 TYR 51 N) -0. 914000000	98. 807000000	19. 841000000	*
153 C (1 TYR 51 C) -0. 908000000	97. 651000000	18. 975000000	*
154 C (1 TYR 51 C) -0. 405000000	96. 559000000	19. 911000000	*
155 O (1 TYR 51 O) 0. 750000000	96. 577000000	20. 346000000	*

途省略

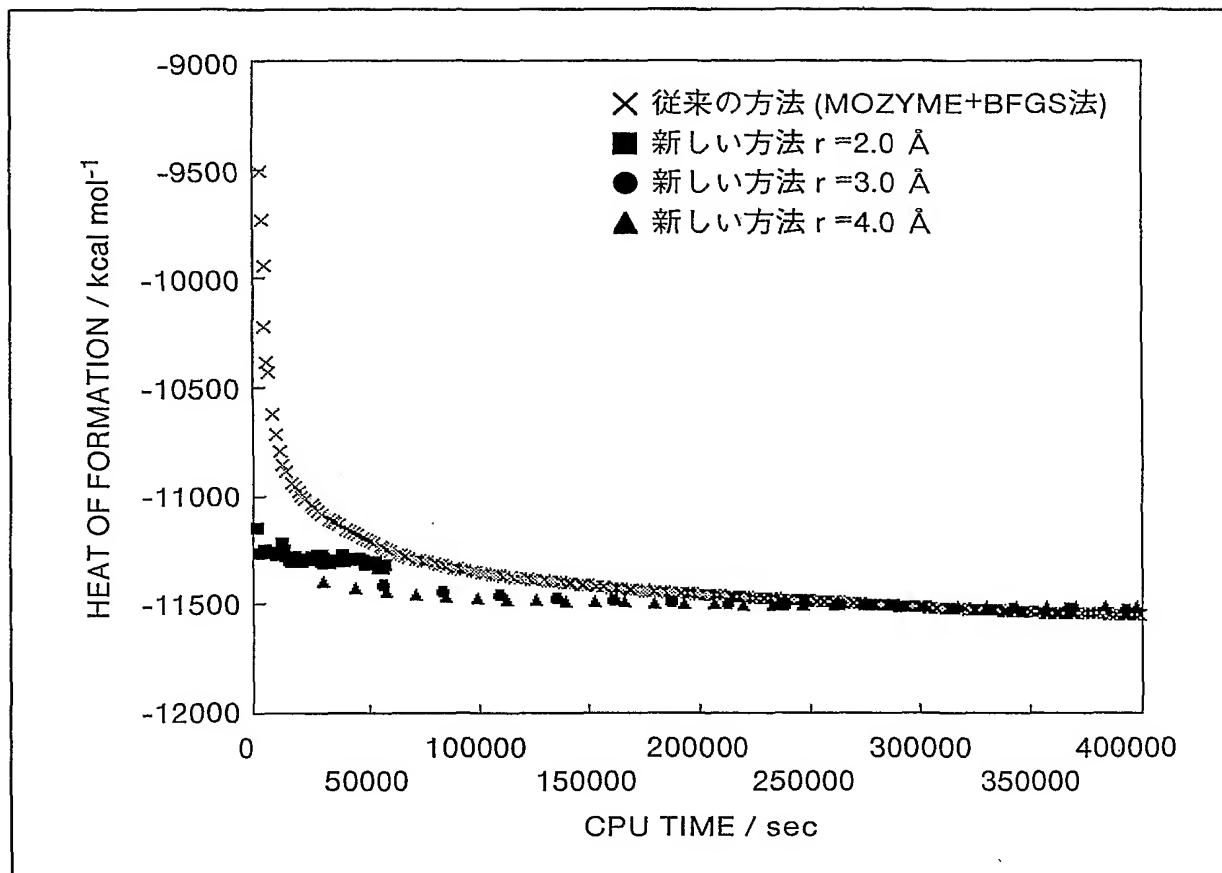
76/77

## 第89図



77/77

## 第90図



## INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP03/06952

## A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl<sup>7</sup> G06F17/30, C07K1/00

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int.Cl<sup>7</sup> G06F17/30, C07K1/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Toroku Jitsuyo Shinan Koho	1994-2003
Kokai Jitsuyo Shinan Koho	1971-2003	Jitsuyo Shinan Toroku Koho	1996-2003

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

JICST FILE (JOIS), WPI, INSPEC (DIALOG)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
E, X	JP 2003-196290 A (Celestar Lexico-Sciences, Inc.), 11 July, 2003 (11.07.03), Claims (Family: none)	1-10
A	JP 5-282381 A (Fujitsu Ltd.), 29 October, 1993 (29.10.93), Claim 1 (Family: none)	1-10
A	FURUTA et al., "Tanpakushitsu no Niji Kozo Yosoku Shien Expert·Expert System", Information Processing Society of Japan Dai 33 Kai (Showa 61 Nen Koki) Zenkoku Taikai Koen Ronbunshu (II), 01 October, 1986 (01.10.86), pages 1197 to 1198	1-10

☒ Further documents are listed in the continuation of Box C.☐ See patent family annex.

* Special categories of cited documents:	"I" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier document but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search  
18 September, 2003 (18.09.03)Date of mailing of the international search report  
14 October, 2003 (14.10.03)Name and mailing address of the ISA/  
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP03/06952

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JP 2000-143554 A (The Institute of Physical and Chemical Research), 23 May, 2000 (23.05.00), Claims (Family: none)	11-24
Y	AKABANE, UMEYAMA, "Yakubutsu-Juyotai Sosui Sogo Sayo", Gendai Kagaku, special extra issue 13, Shin'yaku no Lead Generation -Saishin Drug Design-, 20 November, 1987 (20.11.87), pages 136 to 148, particularly, page 145	25-34
Y	KOMATSU, UMEYAMA, "Yakubutsu-Juyotai Seiden Sogo Sayo", Gendai Kagaku, special extra issue 13, Shin'yaku no Lead Generation -Saishin Drug Design-, 20 November, 1987 (20.11.87), pages 119 to 135, particularly, page 126, right column	25-34
X	WO 93/20525 A (Akiko ITAI), 14 October, 1993 (14.10.93), Page 3, upper left column, line 16 to upper right column, line 2 & EP 633534 A                      & US 5642292 A	35-45
X	SATO et al., "Ab initio pair Kinjiho ni yoru Heiretsu Bunshi Kido Keisan Program ABINIT-MP no Sakusei to Seino Hyoka", Transactions of Information Processing Society of Japan, Vol.41, No.SIG5(HPS1), 15 August, 2000 (15.08.00), page 104 to 112, particularly, p.106	46-55



# INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP03/06952

## Box I Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
  
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
  
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

The technical feature of claims 1-10 is prediction of an interaction portion of a protein by calculation of frustration.

The technical feature of claims 11-24 is prediction of an active portion of a protein or a polypeptide by molecular orbit calculation.

The technical feature of claims 25-34 is determination of an interaction portion of a protein in consideration of the hydrophobic interaction and the electrostatic interaction.

The technical feature of claims 35-45 is prediction of a junction portion by identifying electrostatically unstable amino acid residues on the basis of data on the spatial distances (Continued to extra sheet.)

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
  
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
  
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
  
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest ☐ The additional search fees were accompanied by the applicant's protest.

☒ No protest accompanied the payment of additional search fees.

**INTERNATIONAL SEARCH REPORT**

International application No.

PCT/JP03/06952

Continuation of Box No. II of continuation of first sheet (1)

between amino acid residues of a protein or polypeptide.

The technical feature of claims 46-55 is optimization of a protein structure by electric charge calculation.

Therefore, the number of inventions of claims 1-10, 11-24, 25-34, 35-45, 46-55 is five.

## A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int. Cl<sup>7</sup> G06F17/30, C07K1/00

## B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int. Cl<sup>7</sup> G06F17/30, C07K1/00

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1922-1996年
日本国公開実用新案公報	1971-2003年
日本国登録実用新案公報	1994-2003年
日本国実用新案登録公報	1996-2003年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

JICSTファイル (JOIS), WPI, INSPEC (DIALOG)

## C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
EX	JP 2003-196290 A(セスター・レキシコ・サイエンス株式会社)2003.07.11 特許請求の範囲(ファミリーなし)	1-10
A	JP 5-282381 A(富士通株式会社)1993.10.29, 請求項1(ファミリーなし)	1-10
A	古田 他, 蛋白質の二次構造予測支援エキスパート・エキスパートシステム, 情報処理学会第33回(昭和61年後期)全国大会講演論文集(II), 1986.10.01, p.1197-1198	1-10

☒ C欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

## \* 引用文献のカテゴリー

「A」 特に関連のある文献ではなく、一般的技術水準を示すもの  
「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの  
「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)  
「O」 口頭による開示、使用、展示等に言及する文献  
「P」 国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの

「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの

「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの

「&amp;」 同一パテントファミリー文献

国際調査を完了した日

18.09.03

国際調査報告の発送日

14.10.03

国際調査機関の名称及びあて先

日本国特許庁 (ISA/JP)

郵便番号100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

高瀬 勤



5M

9069

電話番号 03-3581-1101 内線 3597

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
X	JP 2000-143554 A(理化学研究所)2000. 05. 23 特許請求の範囲(ファミリーなし)	11-24
Y	赤羽, 梅山, 薬物-受容体疎水相互作用, 現代化学 増刊13 新薬のリードジェネレーション-最新ドラッグデザイン-, 1987. 11. 20, p. 136-148, 特に, p. 145	25-34
Y	小松, 梅山, 薬物-受容体静電相互作用, 現代化学 増刊13 新薬のリードジェネレーション-最新ドラッグデザイン-, 1987. 11. 20, p. 119-135, 特に, p. 126右欄	25-34
X	WO 93/20525 A(板井昭子)1993. 10. 14 第3頁左上欄第16行-右上欄第2行 &EP 633534 A &US 5642292 A	35-45
X	佐藤 他, Ab initioペア近似法による並列分子軌道計算プログラム ABINIT-MPの作成と性能評価, 情報処理学会論文誌, Vol. 41, No. SIG5 (HPS1)2000. 08. 15, p. 104-112, 特に, p. 106	46-55

## 第Ⅰ欄 請求の範囲の一部の調査ができないときの意見（第1ページの2の続き）

法第8条第3項（PCT17条(2)(a)）の規定により、この国際調査報告は次の理由により請求の範囲の一部について作成しなかった。

1. ☐ 請求の範囲 \_\_\_\_\_ は、この国際調査機関が調査をすることを要しない対象に係るものである。つまり、
2. ☐ 請求の範囲 \_\_\_\_\_ は、有意義な国際調査をすることができる程度まで所定の要件を満たしていない国際出願の部分に係るものである。つまり、
3. ☐ 請求の範囲 \_\_\_\_\_ は、従属請求の範囲であってPCT規則6.4(a)の第2文及び第3文の規定に従って記載されていない。

## 第Ⅱ欄 発明の単一性が欠如しているときの意見（第1ページの3の続き）

次に述べるようにこの国際出願に二以上の発明があるとこの国際調査機関は認めた。

請求の範囲1-10は、フラストレーションの計算によるタンパク質の相互作用部位に予測を技術的特徴とするものである。

請求の範囲11-24は、分子軌道計算によるタンパク質、ポリペプチドの活性部位の予測を技術的特徴とするものである。

請求の範囲25-34は、疎水相互作用、静電相互作用を考慮してタンパク質の相互作用部位を特定することを技術的特徴とするものである。

請求の範囲35-45は、タンパク質、ポリペプチドのアミノ酸残基間の空間距離データ、アミノ酸の電荷から静電的に不安定なアミノ酸残基を特定することによる結合部位の予測を技術的特徴とするものである。

請求の範囲46-55は、電荷計算によるタンパク質構造最適化を技術的特徴とするものである。

したがって、請求の範囲1-10、11-24、25-34、35-45、46-55に記載された発明の数は5である。

1. ☒ 出願人が必要な追加調査手数料をすべて期間内に納付したので、この国際調査報告は、すべての調査可能な請求の範囲について作成した。
2. ☐ 追加調査手数料を要求するまでもなく、すべての調査可能な請求の範囲について調査することができたので、追加調査手数料の納付を求めなかった。
3. ☐ 出願人が必要な追加調査手数料を一部のみしか期間内に納付しなかったため、この国際調査報告は、手数料の納付のあった次の請求の範囲のみについて作成した。
4. ☐ 出願人が必要な追加調査手数料を期間内に納付しなかったため、この国際調査報告は、請求の範囲の最初に記載されている発明に係る次の請求の範囲について作成した。

## 追加調査手数料の異議の申立てに関する注意

- ☐ 追加調査手数料の納付と共に出願人から異議申立てがあった。
- ☒ 追加調査手数料の納付と共に出願人から異議申立てがなかった。